



# An examination of daily CO<sub>2</sub> emissions prediction through a comparative analysis of machine learning, deep learning, and statistical models

Adewole Adetoro Ajala<sup>1,3</sup> · Oluwatosin Lawrence Adeoye<sup>1</sup> · Olawale Moshood Salami<sup>1</sup> · Ayoola Yusuf Jimoh<sup>2,3</sup>

Received: 28 June 2024 / Accepted: 7 December 2024 / Published online: 13 January 2025  
© The Author(s) 2025

## Abstract

Human-induced global warming, primarily attributed to the rise in atmospheric CO<sub>2</sub>, poses a substantial risk to the survival of humanity. While most research focuses on predicting annual CO<sub>2</sub> emissions, which are crucial for setting long-term emission mitigation targets, the precise prediction of daily CO<sub>2</sub> emissions is equally vital for setting short-term targets. This study examines the performance of 14 models in predicting daily CO<sub>2</sub> emissions data from 1/1/2022 to 30/9/2023 across the top four polluting regions (China, India, the USA, and the EU27&UK). The 14 models used in the study include four statistical models (ARMA, ARIMA, SARMA, and SARIMA), three machine learning models (support vector machine (SVM), random forest (RF), and gradient boosting (GB)), and seven deep learning models (artificial neural network (ANN), recurrent neural network variations such as gated recurrent unit (GRU), long short-term memory (LSTM), bidirectional-LSTM (BiLSTM), and three hybrid combinations of CNN-RNN). Performance evaluation employs four metrics ( $R^2$ , MAE, RMSE, and MAPE). The results show that the machine learning (ML) and deep learning (DL) models, with higher  $R^2$  (0.714–0.932) and lower RMSE (0.480–0.247) values, respectively, outperformed the statistical model, which had  $R^2$  (–0.060–0.719) and RMSE (1.695–0.537) values, in predicting daily CO<sub>2</sub> emissions across all four regions. The performance of the ML and DL models was further enhanced by differencing, a technique that improves accuracy by ensuring stationarity and creating additional features and patterns from which the model can learn. Additionally, applying ensemble techniques such as bagging and voting improved the performance of the ML models by approximately 9.6%, whereas hybrid combinations of CNN-RNN enhanced the performance of the RNN models. In summary, the performance of both the ML and DL models was relatively similar. However, due to the high computational requirements associated with DL models, the recommended models for daily CO<sub>2</sub> emission prediction are ML models using the ensemble technique of voting and bagging. This model can assist in accurately forecasting daily emissions, aiding authorities in setting targets for CO<sub>2</sub> emission reduction.

**Keywords** Daily CO<sub>2</sub> emissions · Prediction and forecast · Machine learning model · Deep learning model · Statistical model

## Introduction

Human-induced global warming presents a significant and urgent threat with widespread implications for the environment and society (Kong et al. 2022). This increase in the Earth's average temperature, resulting from the emission of greenhouse gases (nitrous oxide-N<sub>2</sub>O, carbon dioxide-CO<sub>2</sub>, methane-CH<sub>4</sub>, ozone-O<sub>3</sub>, and chlorofluorocarbon-CFC), gives rise to severe consequences, including elevated sea levels, extreme weather events, reduced crop yields, water scarcity, and disturbances to ecosystems (Legg 2021). These outcomes, in turn, pose substantial risks to human well-being. Among the various greenhouse gases released,

Responsible Editor: V.V.S.S. Sarma

✉ Adewole Adetoro Ajala  
woltoaj@gmail.com

- <sup>1</sup> Centre of Excellence for Data Science Artificial Intelligence & Modelling (DAIM), University of Hull, HU6 7RX Hull, United Kingdom
- <sup>2</sup> Department of Geology and Mineral Science, Kwara State University, Malete, P.M.B. 1530, Ilorin, Kwara State, Nigeria
- <sup>3</sup> Research Division, Stalwart Ventures LLC, 5719 Cedonia Avenue, Suite F, Baltimore, MD 21206, USA

CO<sub>2</sub> contributes to approximately 81% of total emissions (Amarpuri et al. 2019; Rehman et al. 2021).

The 2021 global carbon budget revealed that 33% of total CO<sub>2</sub> emissions over the past 70 years occurred after the turn of the millennium (Huang et al. 2023). This substantial release of CO<sub>2</sub> is a major contributor to the issue of global warming (Fakana 2020). The correlation between CO<sub>2</sub> and global warming is rooted in this gas’s ability to trap heat in the Earth’s atmosphere, resulting in a greenhouse effect that elevates the Earth’s temperature. The increase in atmospheric CO<sub>2</sub> levels is attributed primarily to the combustion of fossil fuels, and China, the USA, the EU27&UK, and India account for most of the global emissions (Crippa et al. 2021).

To address the challenge of global warming, the Paris accord, established under the United Nations Framework Convention on Climate Change (UNFCCC), was adopted in 2015. The agreement aims to constrain global warming to either 2 or 1.5 °C by implementing policies that reduce emissions, ultimately achieving net-zero carbon by 2050 (Azevedo et al. 2021). Additionally, in 2021, approximately 200 countries signed the Glasgow Climate Pact (COP26), which aims to encourage governments to “accelerate the development, implementation, and dissemination of

technologies, and the adoption of policies to shift towards a low-carbon emission energy system” (Lennan and Morgera 2022). However, the development of effective policies for setting carbon emission mitigation targets requires the attainment of highly accurate prediction models, thus presenting a new area of challenge. Consequently, there is growing significance in research dedicated to the prediction of CO<sub>2</sub> emissions.

Currently, most of the research is focused on predicting annual CO<sub>2</sub> emissions. While this approach is essential for establishing long-term emission mitigation targets, it has limitations. These limitations arise from its small sample size and its inability to capture daily and monthly fluctuations, which are critical for setting actionable emission reduction policies. In contrast, adopting daily real-time carbon emission prediction allows monitoring of dynamic trends and fluctuations in CO<sub>2</sub> emissions. This approach proves valuable for establishing short-term emission targets, allowing time for effective countermeasures and better control over carbon emissions. Hence, the exploration of daily emission prediction is highly important. Furthermore, an initial exploratory analysis of global CO<sub>2</sub> emissions, as shown in Fig. 1, reveals that more than 61% of total global CO<sub>2</sub> emissions originate from four regions (China, India, the

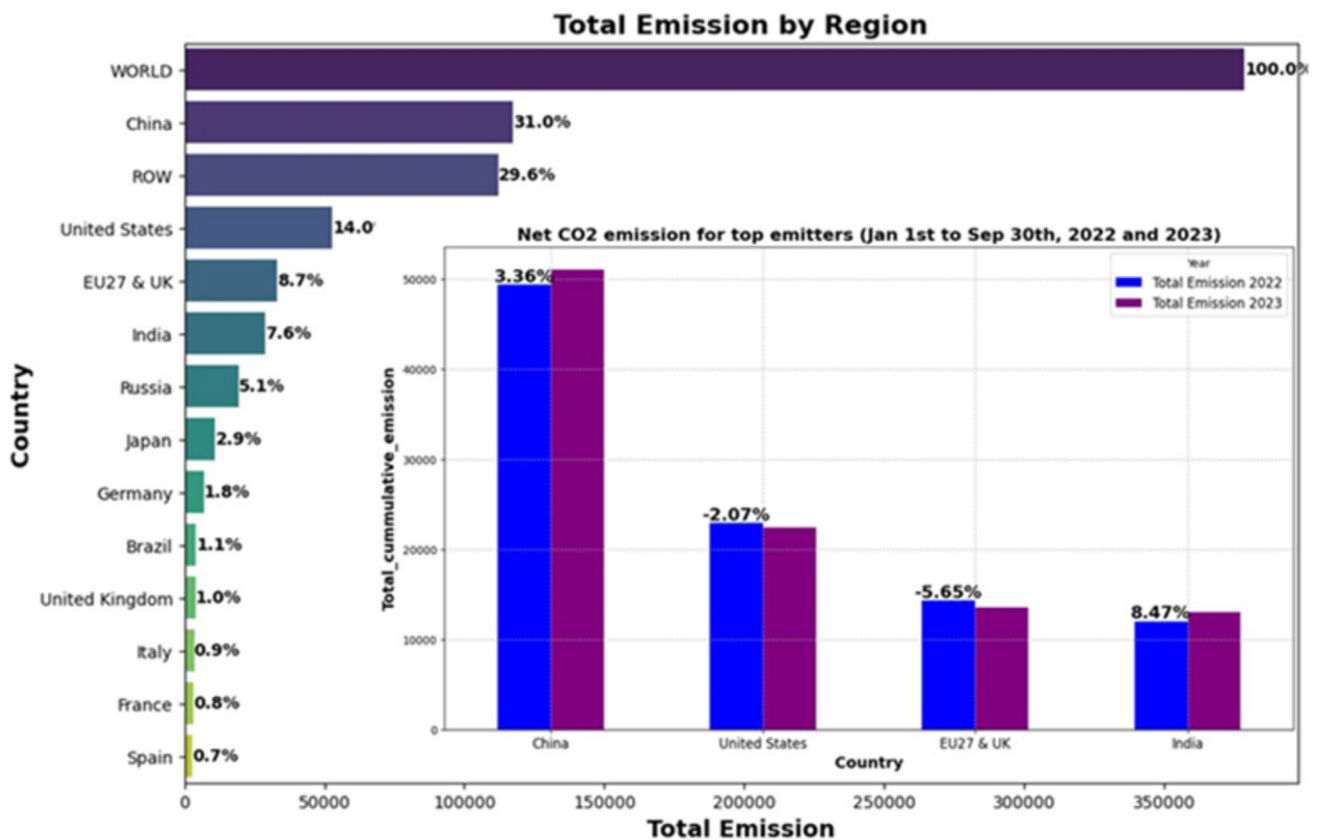


Fig. 1 Global CO<sub>2</sub> emissions between 1/1/2022 to 9/30/2023

USA, and the EU27&UK), with China and India characterised by a net emission increase. Therefore, understanding these regions' current and future daily carbon dioxide emission levels is crucial.

This study investigates the performance of four statistical models (ARMA, ARIMA, SARMA, and SARIMA), three machine learning models (SVM-support vector machine, RF-random forest, and GB-gradient boosting), and seven deep learning models (ANN-artificial neural network, and RNN-recurrent neural network variation of GRU-gated recurrent unit, LSTM-long short-term memory, BiLSTM-bidirectional-LSTM, and three Hybrid combinations of CNN-RNN) in predicting daily CO<sub>2</sub> emission across four regions (China, India, USA, and EU27&UK). The performance of these models is evaluated via four metrics, and a comparative analysis is conducted to identify the best model for daily CO<sub>2</sub> emission prediction. The dataset used for this study comprises daily CO<sub>2</sub> emissions data from 1/1/2022 to 9/30/2023 (638 data points), reflecting post-COVID-19 normal economic activities. This systematic approach ensures that the selected model delivers accurate predictions and demonstrates robustness and reliability. The possible innovations and contributions of this paper are as follows:

- (a) Bridging the Gap in Daily CO<sub>2</sub> emissions prediction: A key contribution of this research is its focus on addressing the critical gap in daily carbon dioxide emissions prediction, which is essential for policymakers in the timely adjustment of actionable short-term emission reduction strategies. Unlike most existing studies that focus on annual carbon dioxide emissions, this study concentrates on the often-overlooked aspect of daily emissions. Doing so fills a significant void in the field and provides valuable insights that can directly inform and enhance policy decision-making.
- (b) Comprehensive model evaluation: This study conducts a thorough evaluation of various prediction models, including statistical, machine learning, and deep learning approaches. The analysis identifies the most effective models for daily CO<sub>2</sub> emission forecasting and provides insights into their strengths and limitations. This evaluation offers valuable guidance for future research in selecting the most appropriate predictive models for daily CO<sub>2</sub> emission forecasting.
- (c) Expansion of geographical scope: Unlike previous studies, which focused primarily on daily emissions in China, this study extends the analysis to include four major polluting regions: China, India, the USA, and the EU27 & UK. By evaluating the performance of prediction models across these diverse regions, the study provides a more comprehensive perspective on the applicability and generalizability of the models, thus enhancing their relevance to a variety of datasets.
- (d) Incorporation of data transformation techniques: This study applied data transformation methods, such as differencing, to improve the accuracy of daily CO<sub>2</sub> emission predictions, particularly in machine learning models that do not rely on the assumption of stationarity. This approach not only improves current predictive capabilities but also offers a foundation for further advancements in the accuracy of carbon emission forecasting in future research.

This study presents new insights backed by empirical evidence, along with practical recommendations aimed at enhancing the accuracy of daily carbon dioxide emission predictions. As a comprehensive analysis, it serves as a valuable resource for researchers, facilitating the development of effective emission reduction strategies and supporting informed environmental policymaking.

The remainder of the paper is organized as follows: the "Literature review" section comprehensively reviews previous work on CO<sub>2</sub> emission forecasting and related studies, establishing the foundation for the present study. The "Methodology" section introduces the data sources, proposed models, and evaluation criteria for assessing the prediction models. This section details the data selection and collection process, outlines the methodologies applied in the prediction models, and describes the metrics used to evaluate model performance. The "Results and discussion" section then presents a detailed evaluation and comparison of each prediction model's performance, examining the results to emphasise their strengths, limitations, and overall efficacy. This section also explores the implications of the findings and their relevance to the research objectives. Finally, the "Conclusion", "Limitations and Future Work" sections summarise the key findings, discuss policy implications, and outline limitations and potential avenues for future research.

## Literature review

Many scientific studies have aimed to increase the precision of CO<sub>2</sub> emission prediction. These studies can be broadly categorised into two groups on the basis of the input data: multivariate and univariate methods. The multivariate approach incorporates multiple factors influencing CO<sub>2</sub> emissions as model inputs. However, this method has a major disadvantage: obtaining complete data for all influencing factors is often challenging in practical applications. As a result, missing data are frequently extrapolated, which may lead to inaccurate predictions or the generation of an erroneous model (Song et al. 2023). Additionally, assessing the individual contribution of each influencing factor is complex. In multivariate models, interactions between variables are often intricate and interdependent, making it difficult to isolate and quantify the impact of specific factors.

Wang et al. (2019) reported that multicollinearity, where predictors are highly correlated, can distort model coefficient interpretations, complicating the identification of individual factor impacts. Prakash and Singh (2024) addressed the missing data issue by employing linear interpolation and mean method techniques to replace missing values in their study on CO<sub>2</sub> emissions from coal power plants. While these techniques help address data gaps, they also introduce uncertainty, potentially affecting model accuracy.

In contrast, the univariate approach relies on readily available historical data, reducing uncertainties in model assumptions and predictions (Ziel and Weron 2018). This method avoids the complexities and challenges associated with missing data in multivariate models, making it preferable when high-quality, comprehensive datasets are unavailable or difficult to obtain. Several recent studies (Giannelos et al. 2023; Kour 2023) have demonstrated that the univariate approach can yield robust predictions with simpler input structures, mitigating risks associated with incomplete or interpolated data. Existing studies on univariate CO<sub>2</sub> emission prediction can be further divided based on the predictive models employed: traditional statistical models and artificial intelligence (AI) models (Mason et al. 2018). Traditional statistical time series models, such as Holt-Winters, ARIMA, and SARIMAX, are commonly applied for short-term CO<sub>2</sub> emission prediction. However, these statistical methods have two significant limitations: (1) they are effective for stationary time series data but struggle with nonlinear cases, and (2) they are not ideal for predicting large sample sizes (Ren et al. 2016; Kong et al. 2022). Despite these limitations, several studies have utilised these models for CO<sub>2</sub> emission prediction. For example, Kour (2023) and Kumari and Singh (2022) effectively applied statistical models to forecast annual CO<sub>2</sub> emissions in South Africa (using ARIMA) and India (using SARIMAX), respectively. However, Li and Zhang (2023) reported that these models performed poorly when applied to daily CO<sub>2</sub> emission data.

Given that daily CO<sub>2</sub> emission data often display nonlinear fluctuations, the linearity assumption inherent in statistical models may lack validity (Peng et al. 2020). Consequently, there has been a shift toward the use of traditional machine learning (ML) and deep learning (DL) models, which are better suited for large datasets and nonlinear data, as highlighted by Zhou et al. (2021). Numerous studies have since focused on predicting CO<sub>2</sub> emissions using traditional ML and DL techniques. These AI models include support vector machines (Adegboye et al. 2024), random forests (Zhang et al. 2023b), gradient boosting regressors (Romeiko et al. 2020), artificial neural networks (ANNs)

(Ağbulut 2022), and recurrent neural networks (RNNs), such as LSTM (Huang et al. 2019) and BiLSTM (Aamir et al. 2022), along with hybrid combinations of convolutional neural networks (CNNs) and RNNs (Amarpuri et al. 2019; Faruque et al. 2022). These models have been applied across various sectors, such as emissions from buildings (Giannelos et al. 2023), CO<sub>2</sub> emissions from transportation (Ağbulut 2022), emissions from coal power plants (Prakash and Singh 2024), CO<sub>2</sub> emissions in the industrial sector (Song et al. 2023), CO<sub>2</sub> emissions from diverse fuel sources (Jeniffer et al. 2023), and energy consumption (Adegboye et al. 2024). A summary of recent studies using the univariate approach for CO<sub>2</sub> emissions prediction is provided in Table 1.

As shown in Table 1, most prior studies have focused primarily on annual CO<sub>2</sub> emissions. While annual predictions are essential for setting long-term mitigation targets, they have limitations, including small sample sizes and an inability to capture daily or monthly fluctuations, which are crucial for actionable short-term policies. In contrast, daily real-time CO<sub>2</sub> predictions allow monitoring of dynamic trends and fluctuations, which is invaluable for setting short-term targets and implementing timely interventions. Consequently, daily emissions prediction represents a critical area for further exploration. Additionally, studies by Magazzino and Mele (2022) and Magazzino et al. (2023) illustrate the potential of combining neural networks with time series decomposition to capture emissions' dynamic behaviour and improve accuracy. Although these studies applied multivariate approaches, they provide insights that complement our strategy of integrating data transformation techniques, such as differencing, to enhance the performance of ML and DL models.

Building on advancements in CO<sub>2</sub> emission prediction, this study introduces several key innovations to address existing gaps. Our research provides a comparative evaluation of 14 models, highlighting variations in performance and identifying the best-performing model. This study broadens the geographical scope by assessing models across China, India, the USA, and the EU27&UK, enhancing the generalizability and applicability of predictive methods. Additionally, by applying data transformation techniques, such as differencing, we improve the accuracy of daily CO<sub>2</sub> emission predictions, especially for the ML and DL models that benefit from stabilised data. By analysing recent daily data from January 2022 to September 2023, reflecting post-COVID-19 economic activity, this study offers empirical evidence and novel insights that meaningfully contribute to the body of literature on short-term CO<sub>2</sub> emission forecasting.

**Table 1** Studies on univariate CO<sub>2</sub> emissions prediction in the past 5 years

Reference literature	Technique	Dataset	Best model	Metrics
Nyoni and Bonga (2019)	ARIMA	1960–2017 (Annual)	ARIMA (2,2,0)	RMSE, MAE & MAPE
Kour (2023)	ARIMA	1960–2016 (Annual)	ARIMA (4,2,3)	RMSE, MAE & MAPE
Giannelos et al. (2023)	ARIMA, Linear Regression, Shallow & Deep Neural Network (NN)	1971–2014 (Annual)	Deep NN	MAPE
Kumari and Singh (2022)	ARIMA, SARIMAX, Holt-Winters, LR, RF, SVM, ANN and LSTM	1980–2019 (Annual)	LSTM	R <sup>2</sup> , MAE, MSE, RMSE, MAPE, MSLE, ME, & MedAE
Geevaretnam et al. (2022)	ARIMA, RF, SVM, ANN	1991–2020 (Annual)	SVM	MAE, RMSE & MAPE
Amarpuri et al. (2019)	CNN-LSTM, Exponential Smoothing	1960–2017 (Annual)	CNN-LSTM	R <sup>2</sup> , MAE, RMSE & MAPE
Huang et al. (2019)	PCA + LSTM, BPNN, GPR	1990–2016 (Annual)	PCA + LSTM	MAPE
Song et al. (2023)	EMD, Elman Neural Network, Univariate Polynomial Regression (TVF-EMD-ENN-UPR)	01/01/2019–31/05/2022 (daily)	TVF-EMD-ENN-UPR	R <sup>2</sup> , RMSE & MAPE
Kong et al. (2022)	ICEEMDAN + Hybrid Prediction Model	01/01/2019–18/06/2021 (daily)	ICEEMDAN -ISSA-ELM	R <sup>2</sup> , RMSE & MAPE
Qiao et al. (2020)	SVM Optimised with Lion Swarm Optimizer and Genetic Algorithm	1965–2017 (annual)	SVM	MAE
Li and Zhang (2023)	GM, ARIMA, SARIMAX, ANN, RF, LSTM	01/01/2020–30/09/2022 (daily)	LSTM	R <sup>2</sup> , MAE, MSE, RMSE & MAPE,

## Methodology

### Dataset

The dataset consists of 638 daily real-time CO<sub>2</sub> emission data points from January 1, 2022, to September 30, 2023, sourced from the carbon monitoring project (<https://carbonmonitor.org>) (Liu et al. 2022). This period was specifically chosen to reflect normal economic activities following the global recovery from the COVID-19 pandemic. The CO<sub>2</sub> emissions, measured in MtCO<sub>2</sub>/day (million tons of CO<sub>2</sub> per day), were calculated as the total contributions from six sectors (domestic and international aviation, ground transportation, power, industrial, and residential) for China, India, the USA, and EU27&UK.

The data for the four regions exhibit high nonlinearity and non-stationarity, as shown in Fig. 2. The descriptive statistics in Table 2 show that China has the highest variability in daily CO<sub>2</sub> emissions, followed by the USA and the EU27&UK, whereas India has the least variability, as indicated by its lower standard deviation. The median values are close to the mean across regions, suggesting relatively symmetrical distributions. However, kurtosis and skewness reveal that China and the EU27&UK have lighter tails,

indicating fewer extreme emission values. Despite India's lower overall variability, its higher kurtosis and skewness suggest occasional spikes in emissions, similar to those in the USA. China had the highest emissions, with a mean of 31 MtCO<sub>2</sub>/day, approximately double that of the USA (14 MtCO<sub>2</sub>/day), the second-highest emitter. The coefficient of variation (CV) was calculated to assess the relative variability in daily CO<sub>2</sub> emissions. China, with the highest absolute emissions, shows stable emissions (CV = 8.03%), whereas the EU27&UK exhibit greater volatility (CV = 18.29%), indicating more significant fluctuations and clarifying the consistency of emissions patterns across regions.

Seasonal variations in CO<sub>2</sub> emissions can be attributed to several factors, including energy demand driven by weather. As shown in Fig. 2, peaks in CO<sub>2</sub> emissions around December and January across all datasets likely reflect increased energy consumption for heating during colder months, whereas warmer periods show higher emissions from air conditioning. Additionally, variations in economic activity, such as shifts in industrial production and transportation patterns, may contribute to temporary spikes or drops in emissions. Thus, the seasonality in CO<sub>2</sub> emissions arises from a combination of weather-related energy demands, economic cycles, and other temporal factors.

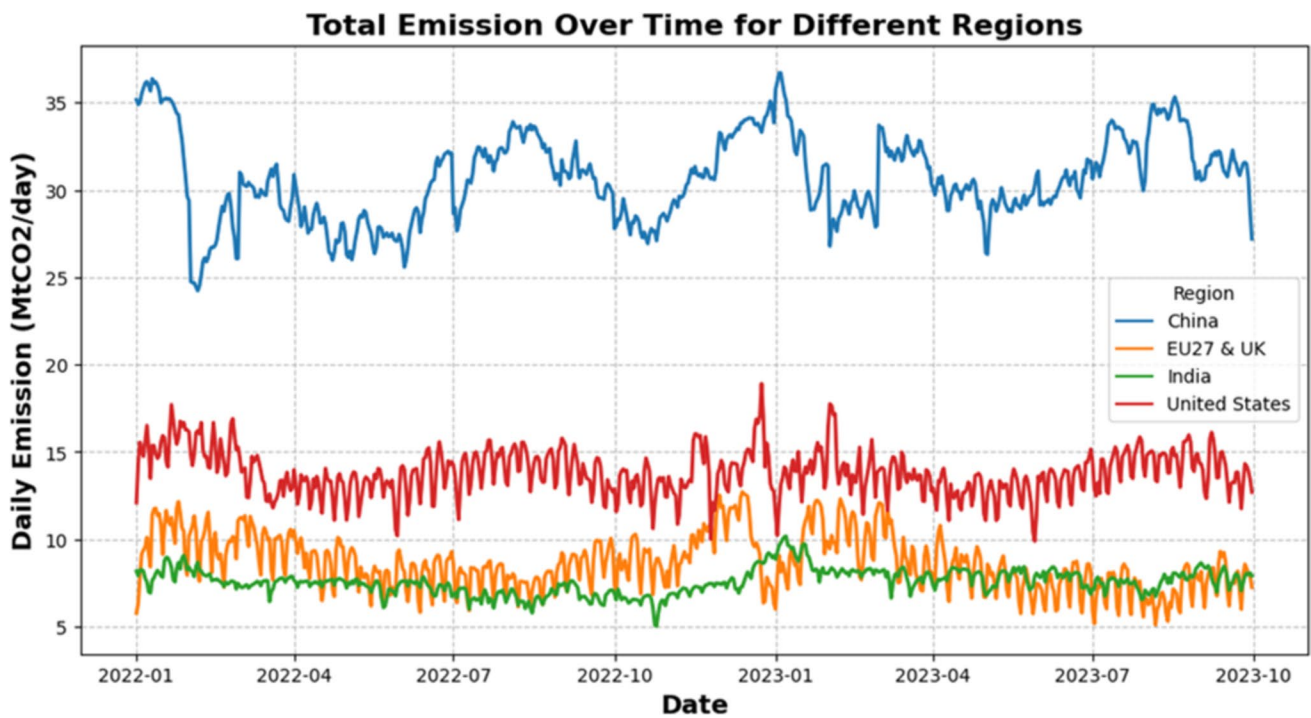


Fig. 2 Daily CO<sub>2</sub> emissions for the four regions

**Table 2** Descriptive statistics of daily CO<sub>2</sub> (MtCO<sub>2</sub>/day) emissions

Country	Count	Mean	Min	Max	Median	Kurtosis	Skewness	Standard-Deviation	CV/%
China	638	30.65	24.23	36.65	30.60	-0.44	0.01	2.46	8.03
USA	638	13.83	9.90	18.91	13.77	0.45	0.15	1.30	9.40
EU27&UK	638	8.64	5.10	12.72	8.46	-0.37	0.35	1.58	18.29
India	638	7.57	5.03	10.18	7.57	0.94	0.26	0.72	9.51

### Data preprocessing

No imputation for missing data was necessary, as the dataset contained no missing values. The main data preprocessing steps are as follows:

#### Stationarity check

The Augmented Dickey-Fuller (ADF) test was utilised to assess the stationarity of the datasets, which is vital for ensuring a consistent mean and standard deviation over time. Stationarity is critical for accurate predictions, particularly in statistical models that rely on this assumption. Although machine learning (ML) and deep learning (DL) models can process non-stationary data, they generally perform better with stationary inputs (Zhou et al. 2021). The ADF test results include both a test statistic and a *p*-value. The test statistic is compared against critical

values to determine whether the null hypothesis (that the data are non-stationary) can be rejected. A more negative test statistic provides stronger evidence against the null hypothesis, indicating that the series is stationary. The *p*-value, derived from this test statistic, quantifies the strength of this evidence, with a *p*-value below 0.05 suggesting stationarity (Ajewole et al. 2020). As shown in Table 3, the China dataset was stationary without any transformation, as indicated by a highly negative test statistic of -3.9787 and a *p*-value of 0.0015. In contrast, the USA, the EU27&UK, and India datasets were non-stationary in their original form, as indicated by their less negative test statistics and *p*-values greater than 0.05. After applying first-order differencing, these datasets achieved stationarity, which was reflected in more negative test statistics and significantly lower *p*-values, well below the 0.05 threshold for rejecting the null hypothesis and assuming stationarity.

**Table 3** ADF test results on the original and transformed data

Dataset	Data transformation	Test statistics	P-value	Conclusion
China	No transformation	−3.9787	0.0015	Stationary
	No transformation	−2.7903	0.0597	Non-stationary
USA	Differencing	−2.9936	0.0350	Stationary
	No transformation	−2.7380	0.0677	Non-stationary
EU27&UK	Differencing	−5.5470	$1.6524 * 10^{-6}$	Stationary
	No transformation	−2.5694	0.0995	Non-stationary
India	Differencing	−5.8816	$3.0687 * 10^{-7}$	Stationary

$P$ -value < 0.05 = Stationary

## Windowing technique

Windowing is a crucial process in transforming time series data into a supervised learning dataset, which is essential for machine learning (ML) and deep learning (DL) models. This technique involves partitioning the time series data into smaller, uniformly sized windows, often called lags. Each window contains a sequence of successive data points that capture the temporal dependencies within the dataset (Moroney 2019). By sliding a fixed-size window across the time series, we can generate input features and corresponding target labels for model training. This sequential iteration allows the model to learn from past values to predict future outcomes, effectively enabling time series forecasting. In this study, a window size of 7 was used. Specifically:

$$\bullet \text{ Input Features} = [X_1, X_2, \dots, X_7], [X_2, X_3, \dots, X_8], \dots, [X_{n-7}, X_{n-6}, \dots, X_n] \quad (1)$$

Here,  $X_i$  represents the data points at each time step, and the window size of 7 captures seven consecutive time points for generating the input features.

$$\bullet \text{ Labels} = [X_8], [X_9], \dots, [X_{n+1}] \quad (2)$$

Each input feature vector consists of seven consecutive time points, such as  $[X_1, X_2, \dots, X_7]$ . The corresponding label is the time point immediately following this window, for example,  $X_8$ . This process is repeated across the entire time series, producing input features and labels for all possible windows in the dataset. For the final window, which includes the last data point  $X_n$ , the label is  $X_{n+1}$ , which follows the last window. This windowing technique helps the model focus on short-term dependencies while allowing it to generalise across different time intervals within the dataset. It is widely used in various applications, including stock market prediction, weather forecasting, and other domains where temporal patterns are critical (Patel et al. 2015; Faruque et al. 2022).

## Models

The study explored four statistical models, three machine learning models, and seven deep learning models to predict and forecast daily CO<sub>2</sub> emissions. The detailed components of each model are explained in the subsequent sections.

### Statistical models

The ARIMA and ARMA models are widely used statistical models that combine autoregression (AR) and moving average (MA) techniques. In ARIMA, “I” signifies integration, indicating the differencing order required to make the time series stationary (Kumari and Singh 2022). The parameters for AR, I, and MA are represented as (p, d, q), where p represents past values for predictions, d is the order of differencing ensuring stationarity, and q is the lag number of prediction errors used to enhance the current timestamp (Kour 2023). An ARIMA model is derived when these parameters are greater than 0. When d equals 0, it corresponds to the ARMA model. The ARIMA model can be expressed mathematically as follows.

$$\varphi_p(B)(1 - B)^d y_t = \theta_q(B)\varepsilon_t \quad (3)$$

where  $\varphi_p$  = autoregressive coefficient; B = backshift operator; d = order of differencing;  $y_t$  = original time series data;  $\theta_q$  = moving average; and  $\varepsilon_t$  = white noise or error term (Cho et al. 1995). The main advantage of ARIMA and ARMA models is their simplicity and effectiveness in capturing linear patterns. However, these models have certain limitations. One drawback is that they require the original or differenced time series data to be stationary. Additionally, they may not perform well with nonlinear data, as they are more attuned to linear patterns (Li and Zhang 2023).

Introducing seasonality into the modelling process can enhance the model’s predictive performance when dealing with seasonal data (Kumari and Singh 2022). Unlike ARMA

and ARIMA, SARIMA incorporates additional parameters specifically tailored for capturing seasonal patterns. These parameters include the seasonal autoregressive parameter (AR)  $P$ , seasonal differencing  $D$ , and seasonal moving average (MA)  $Q$ , denoted as  $P$ ,  $D$ ,  $Q$ , and  $s$ , respectively. Collectively, the SARIMA model integrates both seasonal and nonseasonal components, resulting in a total of  $p$ ,  $d$ ,  $q$ ,  $P$ ,  $D$ ,  $Q$ , and  $s$  parameters, represented as SARIMA  $(p, d, q) \times (P, D, Q, s)$ . The mathematical expression of the SARIMA model is as follows:

$$\Phi_p(B^s)\varphi_p(B)(1 - B^s)^D(1 - B)^d y_t = \Theta_Q(B^s)\theta_q(B)\varepsilon_t \quad (4)$$

where  $B^s$  = the seasonal reverse potential;  $\Phi_p$  = the seasonal AR coefficient;  $\Theta_Q$  = the seasonal MA coefficient; and  $(1 - B^s)^D$  = the seasonal differencing of the order  $D$  (Lee and Han 2020).

The model selection process for these statistical models typically begins with ensuring stationarity in the time series data by estimating the necessary differencing orders ( $d$  and  $D$ ). Once stationarity is achieved, the appropriate model parameters ( $p$ ,  $d$ ,  $q$  for ARIMA/ARMA;  $P$ ,  $D$ ,  $Q$ ,  $s$  for SARIMA) are identified via visual tools such as PACF and ACF plots (Sen et al. 2016). These parameters are then fine-tuned through hyperparameter tuning to obtain the optimal configuration that minimises the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), ensuring the most effective balance between model complexity and goodness of fit. After fitting the model to the training data, residual analysis is performed to detect any systematic patterns or errors in the residuals, ensuring that the model's errors are random and unbiased. Model validation is then conducted by comparing the model's predictions against actual test data and assessing performance metrics such as the R-squared value. Once validated, the model is used to make predictions on new, unseen data, providing insights into future trends or patterns in the time series.

### Machine learning models

This study employed support vector regressor (SVR), ensemble methods such as random forest (RF), and gradient boost regressors (GB) because of their successful application in CO<sub>2</sub> emission prediction, as reported by several researchers (Kumari and Singh 2022; Ağbulut 2022; Romeiko et al. 2020).

#### a) Support Vector Regressor (SVR)

SVR is based on statistical learning theory and works by mapping the input data into a higher-dimensional feature

space using kernel functions (Vapnik 1998). In this transformed space, it finds the optimal hyperplane that separates the data while maximising the margin of tolerance or error, effectively balancing complexity and accuracy. This hyperplane acts as a predictive model, estimating the output for new data. The basic equation for the SVM can be expressed as:

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) + b \quad (5)$$

where  $\alpha_i$  represents the model parameters (Lagrange multipliers),  $K(x_i, x)$  represents the kernel function that transforms the data into higher dimensions, and  $b$  is the bias term (Mohammadi et al. 2015).

The most commonly used kernel functions are the linear, polynomial, and radial basis functions (RBF). The choice of the kernel function plays a critical role in capturing linear and nonlinear relationships in daily CO<sub>2</sub> emission patterns. Linear and RBF kernels are used in this study because they are relatively simple to implement and involve tuning fewer parameters than more complex kernels while providing effective predictive performance (Zhang et al. 2021). The linear kernel works well with linearly separable data, whereas the RBF kernel captures nonlinear relationships efficiently.

The performance of the SVR is optimised by tuning three key parameters: the kernel function, epsilon ( $\epsilon$ ), and the regularisation parameter  $C$ . The epsilon value defines a margin of tolerance within which predictions are considered accurate, which helps to ensure that the model generalises well to unseen data. The  $C$  parameter controls the trade-off between minimising errors on the training data and maintaining a smooth model; higher values of  $C$  lead to fewer errors but may risk overfitting, whereas lower values encourage a simpler model that may tolerate some errors (Kleynhans et al. 2017).

For CO<sub>2</sub> emission forecasting, SVR offers several advantages. It can model complex and nonlinear patterns in the data, making it robust for handling real-world time series data that may exhibit irregularities and fluctuations. By adjusting the kernel, epsilon, and  $C$  parameters, SVR can effectively learn from historical emission data and make reliable future predictions, providing insights into long-term CO<sub>2</sub> emission trends (Wang et al. 2023).

#### b) Ensemble model

In addition to SVR, ensemble methods such as random forest regressor and gradient boosting regressor also provide significant advantages for CO<sub>2</sub> emission prediction (Wang



**Table 4** Tested parameters for the machine learning models

SVR	RF	GB
C: {0.1,1,10}	n-estimators: {50,100,200}	n-estimators: {50,100,200}
Kernel: {linear, rbf}	max-depth: {range (1,6)}	Learning rate: {0.01, 0.1, 0.2}
Degree: {2,3,4}	Min-samples-split: {2, 5,10}	max-depth: {2,3,4}

et al. 2023; Zhang et al. 2023b). Ensemble learning leverages the combined strength of multiple machine learning models trained on the same dataset. Consequently, compared with individual machine learning models, ensemble models can increase forecast accuracy.

**Random forest (RF)** The concept of random forests was introduced by Breiman (2001). Random forest is an ensemble learning method that uses decision trees as its basic units and combines them with the bagging technique to enhance performance. During training, the bootstrap resampling technique is applied to randomly select  $k$  samples from the original dataset, constructing  $k$  weak decision trees (Breiman 2001). Each decision tree grows independently, without constraints or pruning, and remains uncorrelated with the others. The model's final output is determined by averaging the predictions (for regression) or through majority voting (for classification) (Tang and Zhang 2013). Averaging predictions across multiple models serves to reduce variance and enhance the stability of the trees' predictive performance. Random forests excel in capturing complex nonlinear relationships between inputs and outputs, making them particularly effective for regression tasks (Fan 2017). The fundamental equation for ensemble prediction is as follows: for any given sample  $X$  with  $P$  sub-models, each sub-model will generate its own prediction value. Let  $Y_k$  represent the predicted value of the  $k^{\text{th}}$  sub-model; the overall model  $Y_E$  will produce the final result by averaging these predictions (Yoon 2021).

$$Y_E = \frac{1}{P} \sum_{k=1}^P Y_k \quad (6)$$

A major issue with the random forest model is overfitting. This occurs when a model fits the training data almost perfectly (performs well on the training set) but struggles to generalise to new, unseen data. Fully developed trees can lead to overfitting by capturing noise in addition to the underlying patterns in the training data. To address this issue, several adjustments can be made to the random forest model. Pruning the trees by reducing the maximum depth or limiting the number of nodes can help prevent overfitting. This study employed hyperparameter tuning to enhance the model's performance. The key parameters considered included the maximum depth of the trees, the number of estimators, and the minimum number of samples required to split an internal node, as presented in Table 4.

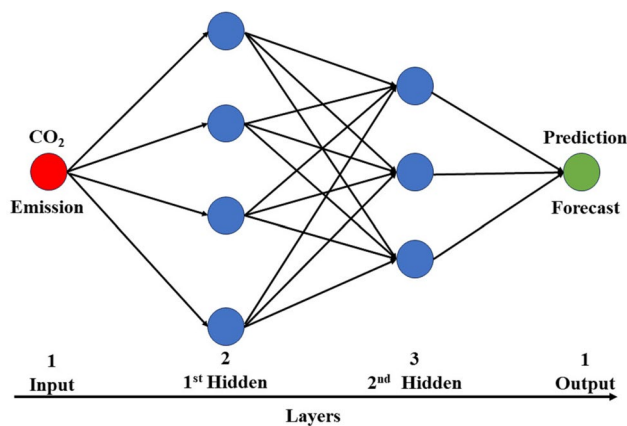
**Gradient boosting regressor (GB)** In contrast, the gradient-boosting regressor builds trees sequentially, with each subsequent tree aiming to rectify the errors of its predecessors (Friedman 2001). The model begins by constructing the first regression tree and then iteratively builds subsequent trees, splitting the data into smaller groups at each step. After each tree is built, the model evaluates the errors and trains the next tree to address those mistakes. This process continues until either the specified number of trees is reached or no further improvement in fit is possible. To prevent overfitting, gradient boosting uses a learning rate to scale the contribution of each new tree (Yoon 2021). A lower learning rate improves the model's ability to generalise, reducing the likelihood of overfitting by controlling how much each tree contributes to the final model. The key parameters tuned to improve the gradient boosting model's performance include the trees' maximum depth, the number of estimators, and the learning rate.

Overall, the grid search method was used to select the optimal parameters needed to enhance the performance of the RF, GB, and SVR models, as outlined in Table 4. Additionally, ensemble methods such as voting, bagging, and stacking were applied to further improve the models' accuracy.

## Deep learning models

### (a) Artificial neural networks (ANN)

Artificial neural networks are well-suited for capturing the complexities of nonlinear data, and their application in modelling CO<sub>2</sub> emission data has been demonstrated by Ahmed et al. (2023). The strength of ANNs is their capacity for self-learning, which makes them particularly useful for managing nonlinear datasets such as daily emission data (Tümer and Akkuş 2018). A typical ANN consists of three layers, namely, the input-hidden-output layers, as shown in Fig. 3 (Ağbulut 2022). The hidden layer functions as the processing layer, enabling the model to learn patterns within the data. This layer consists of neurons (fundamental processing units) that receive input data from the input layer. The hidden layers adapt and learn based on the selected learning method for each processing layer, subsequently transmitting the acquired knowledge



**Fig. 3** Schematic representation of the ANN model (modified after Ağbulut 2022)

to the final layer, known as the output layer (Haider et al. 2022). The model parameters used in this research are shown in Table 5, and the mathematical expression for an ANN is presented as follows (Guo et al. 2021):

$$A_n = \sum_{j=1}^n (w_j \cdot I_j) + b \tag{7}$$

Here,  $n$  represents the number of inputs,  $w_j$ , and  $b$  correspond to the weight and bias, respectively, and  $I_j$  represents the input and  $A_n$  denotes the output of the ANN.

(b) Recurrent neural network (RNN)

Traditional convolutional neural networks (CNNs) are not designed to handle temporal dynamics, as they lack mechanisms for processing and propagating information across time steps. In contrast, recurrent neural networks (RNNs) are specifically designed to incorporate information from previous time steps into their current outputs (Ding et al. 2022). For daily CO<sub>2</sub> emission prediction, recognising the

time-dependent nature of the data is crucial, as it reveals a correlation between daily emission values over time. RNNs are particularly suitable for time-series modelling because of their loops in processing units, which facilitate the handling of sequential data. However, simple RNNs often struggle with capturing long-term dependencies, leading to issues such as vanishing and exploding gradient problems during training (Haider et al. 2022). This limitation can be mitigated by employing advanced RNN variants, such as gated recurrent units (GRUs), Long short-term memory (LSTM) networks, and bidirectional-LSTM (BiLSTM) networks.

LSTM consists of a memory cell denoted by  $C_t$  and three gates: an input gate, a forget gate, and an output gate, which help retain and discard information as needed, effectively capturing long-term relationships (Hochreiter and Schmidhuber 1997). The architectural structure of the LSTM is presented in Fig. 4a, where the input, output, and forget gates are typically represented by  $i_t$ ,  $O_t$ , and  $f_t$  respectively. The input gate receives daily CO<sub>2</sub> emission data, which the forget gate processes by merging with past hidden states. The forget gate determines which information from prior time steps should be retained or discarded. The concatenated input is passed through a non-linear function, which stochastically updates values based on the current data. The output gate generates the predicted CO<sub>2</sub> emissions, whereas a sigmoid layer, tanh layer, and pointwise operations such as summation and multiplication assist in the system's computations. The computing equation for an LSTM unit is presented as follows (Ding et al. 2022):

a) Input:

$$g_t = \tanh(W_{ig}x_t + b_{ig} + W_{hc}h_{t-1} + b_{hg}) \tag{8}$$

b) The Gate Status:

$$input\ gate(i_t) = \sigma(W_{ij}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \tag{9}$$

$$forget\ gate(f_t) = \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \tag{10}$$

$$Output\ gate(O_t) = \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \tag{11}$$

c) Memory Cell:

$$C_t = f_t \times C_{t-1} + i_t \times g_t \tag{12}$$

d) The output:

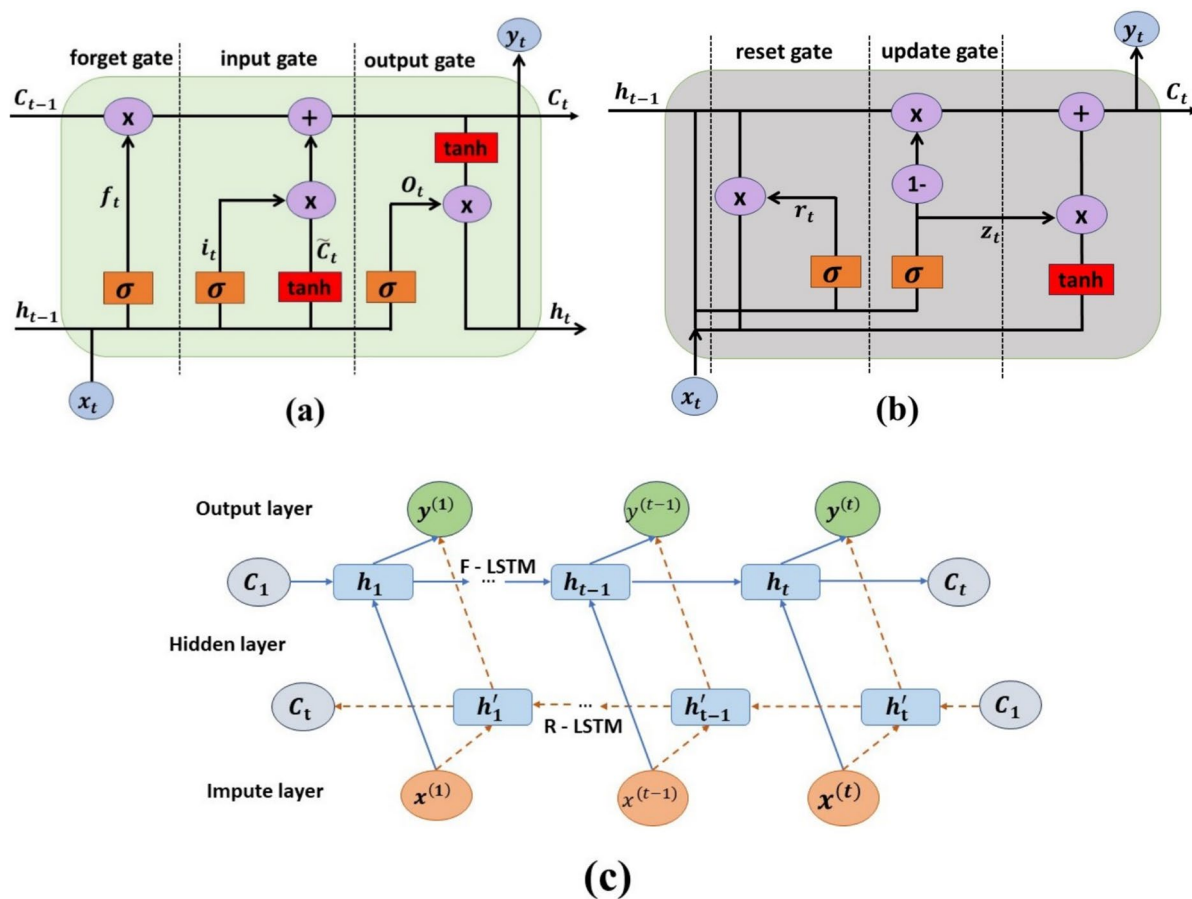
$$h_t = O_t + \tanh(C_t) \tag{13}$$

where  $W$  and  $b$  represent the weight and bias matrices, respectively.  $h_{t-1}$  represents the hidden state (unknown vectors) from the previous time step.  $x_t$  denotes the input at the current time step. These components collectively manage the flow of information within the LSTM,

**Table 5** Model parameters of the deep learning models

Parameters	ANN	GRU	LSTM	BiLSTM	CNN-RNN
Kernal size	Na	Na	Na	Na	varied
Filter size	Na	Na	Na	Na	varied
Conv1D	Na	Na	Na	Na	1
Drop out	0.2	0.2	Varied	0.2	Varied
Number layer	3	3	3	3	4
Learning rate	0.001	0.001	0.001	0.001	varied
Optimizer	Adam	Adam	Adam	Adam	Adam
Batch size	32	32	32	32	32
No of Epochs	200	200	200	200	200
Error monitored	MSE	MSE	MSE	MSE	MSE

Na, not applicable



**Fig. 4** Schematic representation of the basin unit structure of the (a) LSTM and (b) GRU (c) BiLSTM (modified after Ahem et al. 2023; Ding et al. 2022)

enabling it to capture relationships between various data points during prediction and retain critical information. As a result, LSTMs often produce more accurate results when handling time-series data predictions.

The GRU, a simpler variant of the RNN, is similar to the LSTM but uses two gates: the reset gate (combining the forget and input gates) and the update gate, which are typically denoted by  $r_t$  and  $z_t$ , respectively, as shown in Fig. 4b (Hochreiter and Schmidhuber 1997). Reducing the number of gates contributes to GRUs’ faster training times and lower computational complexity. The key equations in the GRU are as follows (Ding et al. 2022):

a) Update gate:

$$z_t = \sigma(W_z x_t + R_z h_{t-1} + b_z) \tag{14}$$

The update gate provides adaptive regulation of information passing through the hidden unit, determining how much of the previous hidden state  $h_{t-1}$  should

be retained and combined with the incoming data for information updates.

b) Reset gate:

$$r_t = \sigma(W_r x_t + R_r h_{t-1} + b_r) \tag{15}$$

The reset gate controls how much of the past information should be reset or discarded.

c) Candidate hidden state:

$$h_t' = \tanh(W_u x_t + R_u (r_t \times h_{t-1}) + b_u) \tag{16}$$

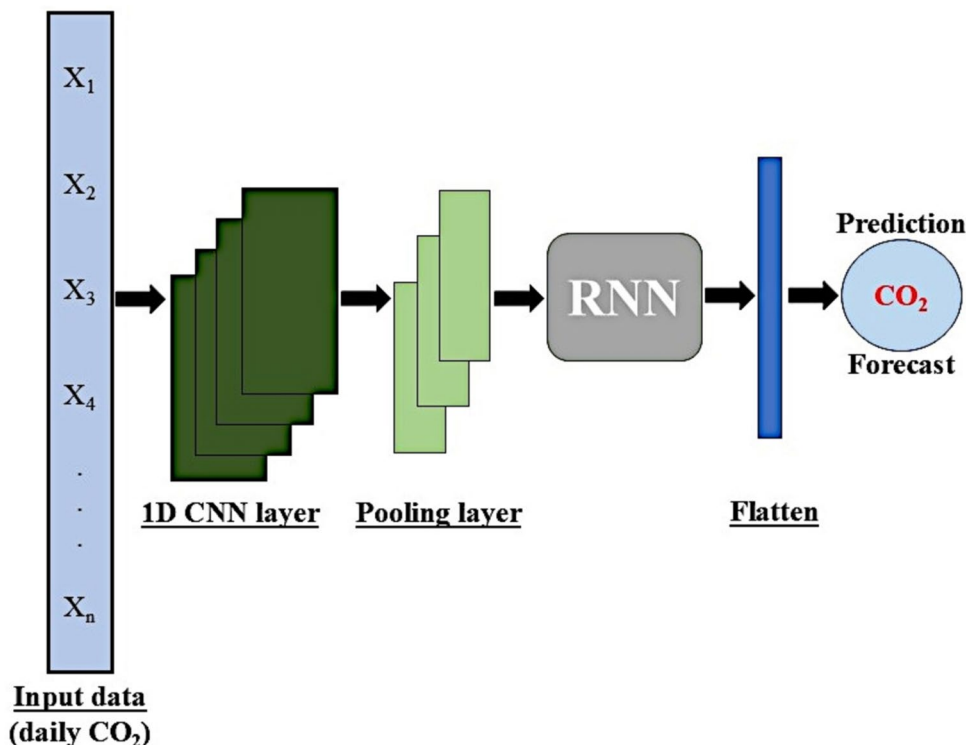
The candidate’s hidden state  $h_t'$  is computed via the reset gate-modulated previous hidden state and the current input  $x_t$

d) Final hidden state:

$$h_t = z_t \times h_t' + (1 - z_t) \times h_{t-1} \tag{17}$$

The final hidden state  $h_t$  is a combination of the candidate’s hidden state  $h_t'$  and the previous hidden state  $h_{t-1}$ , weighted by the update gate  $z_t$ .

**Fig. 5** Architecture of the hybrid CNN-RNN model (modified after Faruque et al. 2022)



Compared with LSTM networks, which have three gates (input, forget, and output) and a separate memory cell, the simpler structure of a GRU, with fewer gates, results in lower computational complexity. This simplicity allows GRUs to train faster while still maintaining effective performance in capturing temporal dependencies, which can be beneficial for daily CO<sub>2</sub> emission forecasting.

Compared with the traditional unidirectional LSTM, the BiLSTM developed by Schmidhuber (2015) consists of two LSTMs with outputs in opposite directions, capturing information from both past and future time steps (seasonality), thereby improving the model's dependence and enhancing the model's overall forecasting accuracy. For BiLSTM, the hidden layer  $h_t$  consists of both forward  $h_t^{\rightarrow}$  and backward  $h_t^{\leftarrow}$  LSTM units, which are expressed as follows:

$$h_t = h_t^{\rightarrow} \oplus h_t^{\leftarrow} \tag{18}$$

where  $\oplus$  denotes the element-wise summation of the forward and backward output components, and the BiLSTM network structure is illustrated in Fig. 4c. Numerous researchers have used these three RNN variations for time series forecasting (Zhang et al. 2023a; Li and Zhang 2023; Huang et al. 2019).

Additionally, we explored a hybrid CNN-RNN model by incorporating a single convolutional layer into each RNN architecture. This integration allows the model to extract both spatial and temporal features simultaneously. The convolutional and pooling layers extract spatial features, whereas the RNN layers capture the time

series information. This approach has shown potential for improving the forecast accuracy of time series models, as reported by numerous researchers (Faruque et al. 2022; Amarpuri et al. 2019).

In the hybrid model, the input daily CO<sub>2</sub> emission data are first fed into the CNN layer, where a convolution operation is applied to extract the characteristics of the complex, non-linear interactions in the input data. The new feature matrix generated after the convolution operation is then passed to the pooling layer, which reduces the dimensionality of the feature map and helps prevent overfitting. This feature map is used as the input for the RNN layer. Finally, a fully connected layer is placed at the end to make the prediction, as shown in Fig. 5 below. The hybrid model uses a ReLU activation function for the input layers, whereas the output layer uses a linear activation function. The parameters of the deep learning models used in this study are listed in Table 5.

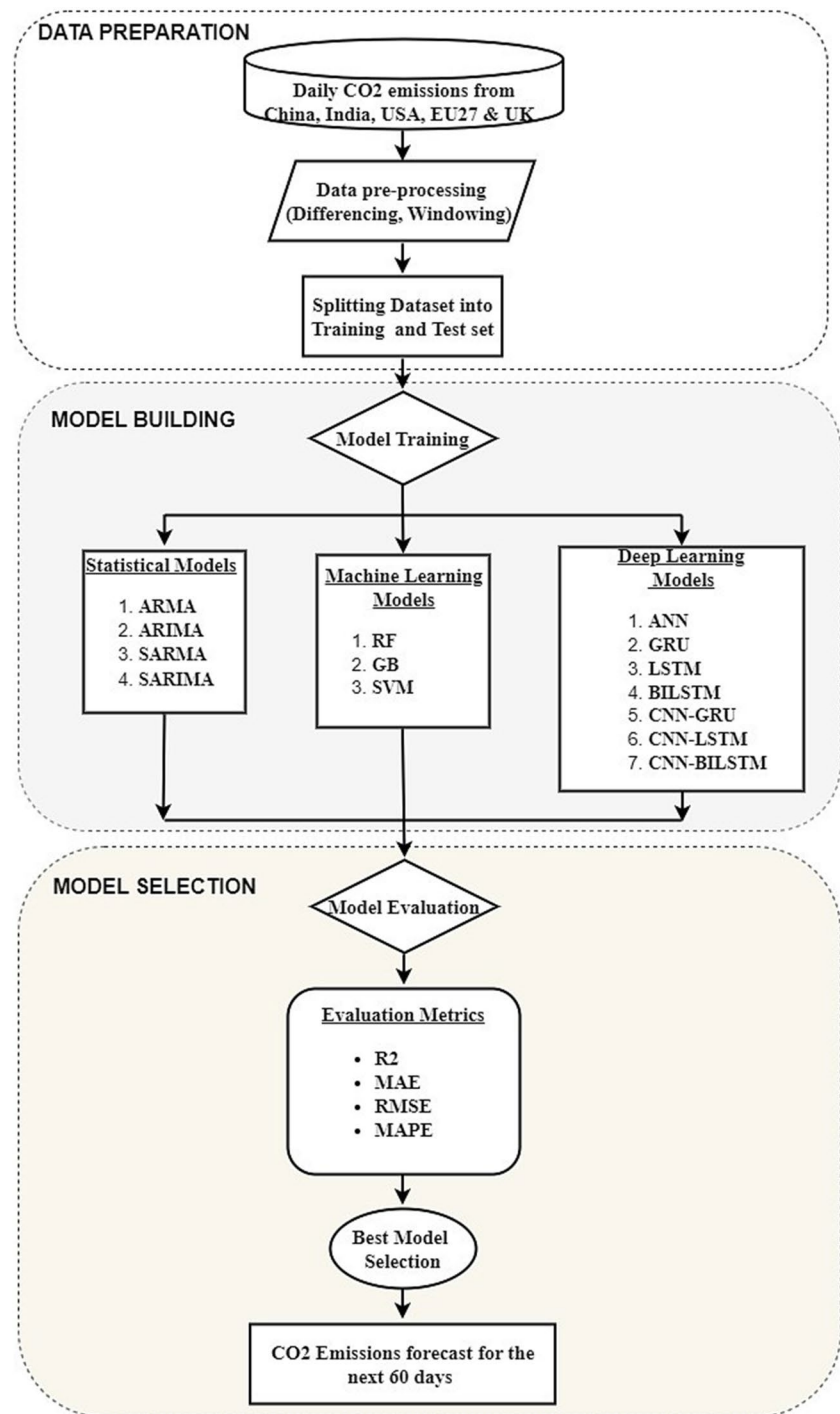
**Research workflow**

The research workflow is divided into three key phases: data preparation, model building, and model selection, as illustrated in Fig. 6.

- (1) Data preparation:

Daily univariate CO<sub>2</sub> emission data for China, India, the USA, and the EU27&UK were collected. To make the data

Fig. 6 Research workflow



suitable for time series models, differencing was applied to ensure stationarity, which is essential for models such as ARIMA and SARIMA, which require stationary input data. Windowing techniques were employed to prepare the data for the ML and DL models, ensuring that temporal

dependencies were captured. The dataset was split into training (80%) and testing (20%) sets to ensure robust model evaluation. The entire data preparation process was carried out via Python's Panda library.

(2) Model building:

The conceptual design of this study involves comparing various models to determine which best fits the CO<sub>2</sub> emission data across different regions. The choice of models was informed by their theoretical foundation and empirical success in previous CO<sub>2</sub> emission forecasting studies (Shabani et al. 2021; Kumari and Singh 2022; Faruque et al. 2022). Specifically:

- Statistical models: ARMA, ARIMA, SARMA, and SARIMA were selected on the basis of their success in capturing linear trends and seasonal patterns in stationary time series data. Studies such as those of Kour (2023) and Kumari and Singh (2022) have demonstrated the efficacy of these models in CO<sub>2</sub> emission forecasting.
- The machine learning models SVM, random forest (RF), and gradient boosting (GB) were chosen for their flexibility in handling nonlinear relationships and high-dimensional data, as demonstrated by Adegboye et al. (2024), Zhang et al. (2023b), and Romeiko et al. (2020). These models do not rely on stationarity assumptions and can capture interactions between variables that traditional statistical models might overlook.
- The deep learning models ANN, GRU, LSTM, and BiLSTM were used to model long-term dependencies in the data. These models have been shown in studies such as those of Shabani et al. (2021), Huang et al. (2019), and Aamir et al. (2022) to be effective at capturing nonlinear, non-stationary patterns in CO<sub>2</sub> emission datasets. Hybrid models that combines CNN and RNN components have also been explored to enhance the ability to capture complex temporal relationships, as Faruque et al. (2022) demonstrated. All modelling and analyses were

performed using Python libraries such as scikit-learn (for the ML models), stats-models (for the statistical models), and TensorFlow/Keras (for the DL models). Hyperparameter tuning was conducted via a grid search to identify optimal model configurations.

(3) Model selection:

Model selection was based on a comprehensive evaluation using metrics such as the RMSE, MAE, MAPE, and R<sup>2</sup>, as recommended by Li and Zhang (2023). These metrics provide insights into both the scale and direction of prediction errors, offering a well-rounded assessment of model performance. Model comparison and visualisation were performed using Python’s matplotlib library. The final model selection was driven by both empirical performance and theoretical suitability, with the selected model used to forecast CO<sub>2</sub> emissions for 60 days across the four major regions. This approach balances traditional statistical theory with modern ML/DL techniques to ensure that the most appropriate model is chosen for accurate daily CO<sub>2</sub> emission forecasting.

**Performance evaluation**

Model performance was evaluated via metrics such as the coefficient of determination (R<sup>2</sup>) and error metrics such as the root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). Smaller values of the error metrics and a higher R<sup>2</sup> value indicate better model performance. Table 6 provides a concise overview of the various evaluation metrics employed in this study.

**Table 6** Summary of the evaluation metrics used in this study

Metrics	Equation	Description
RMSE	$= \sqrt{\sum_{i=1}^n \frac{(x_i - y_i)^2}{n}}$	It is the root of the MSE and has the same unit as the target variable, making it more interpretable. The closer the result is to zero, the better the performance of the model (Ağbulut et al. 2021)
MAE	$= \frac{\sum_{i=1}^n (x_i - y_i)}{n}$	It represents the sum of the absolute deviation of the prediction from the actual value, and a smaller value is desirable (Kumari and Singh 2022)
MAPE	$= \frac{100\%}{n} \sum_{i=1}^n \left  \frac{y_i - x_i}{y_i} \right $	It indicates the model’s average prediction accuracy, with a smaller value being desirable (Song et al. 2023)
R <sup>2</sup>	$= 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$	It varies from 0 to 1, denoting the level of correspondence between the predicted and actual values (Song et al. 2023). The more the value approaches 1, the better the performance of the model
	$y_i, x_i =$	<i>Actual and predicted values</i>

## Results and discussion

### Results

#### Statistical models

On the basis of the performance metrics presented in Table 7, it is evident that the ARMA model is the best statistical model, exhibiting lower MAE, RMSE, and MAPE values, along with a higher  $R^2$  value across regions. However, the incorporation of differencing using ARIMA and SARIMA models resulted in a decrease in the overall model performance. For example, the USA dataset's  $R^2$  decreased from 0.412 to 0.205 (ARIMA) and 0.321 (SARIMA). Similar observations occurred across all the other regions. This observation aligns with the findings of Kumari and Singh (2022), who emphasise that while differencing aids in achieving stationarity by removing trends and seasonality, it may also lead to a loss of valuable information, resulting in inaccurate predictions.

The poor performance of ARIMA and SARIMA models in datasets such as those from China and India, as reflected by negative  $R^2$  values, underscores the limitations of these models in capturing the high variability and nonlinear patterns present in CO<sub>2</sub> emission data. While differencing helps achieve stationarity, it may inadvertently remove significant information, particularly in highly volatile datasets such as

those from China. Li and Zhang (2023) similarly highlighted that traditional models such as ARIMA and SARIMA struggle with nonlinear, volatile datasets.

In summary, while ARMA performed moderately well in regions such as the USA and the EU27&UK, with  $R^2$  values of 0.412 and 0.719 respectively, its overall performance was suboptimal for the China and India datasets, where the  $R^2$  values were mostly negative ( $R^2 \leq 0.068$ ), indicating that the predicted values significantly differ from the actual values, as illustrated in Fig. 7. These models' limitations in handling nonlinearity and volatility make them less suitable for forecasting daily CO<sub>2</sub> emissions.

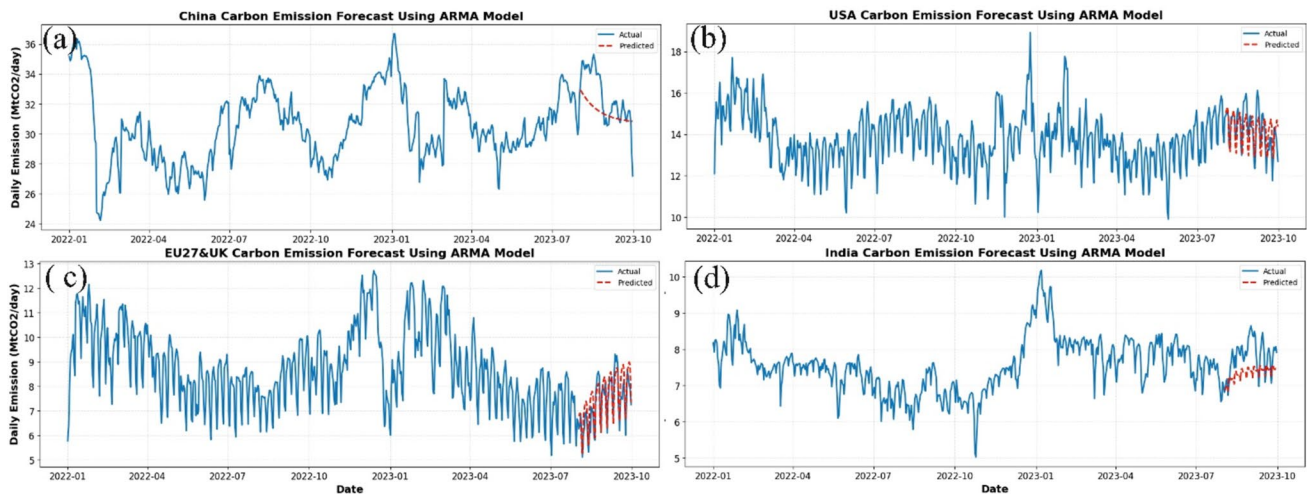
#### Machine learning models

As demonstrated in Table 8, the performance of the ML models significantly improved with differencing compared with that of the statistical models. For example, differencing increased the  $R^2$  values of the RF model from 0.899 to 0.918, 0.585 to 0.804, 0.644 to 0.852, and 0.663 to 0.720 for China, the USA, the EU27&UK, and India datasets, respectively. A similar improvement trend was observed for all the other evaluation metrics (MAE, RMSE, and MAPE) across the four regions. While ML models such as RF, SVM, and GB do not inherently require stationarity, they can be sensitive to non-stationary data, leading to the model capturing

**Table 7** Statistical models performance evaluation

Region	Model	MAE	RMSE	$R^2$	MAPE
China	<b>ARMA (1,0,1)</b>	<b>1.351</b>	<b>1.695</b>	<b>0.068</b>	<b>4.072</b>
	ARIMA (2,1,1)	1.617	1.824	-0.079	5.074
	SARIMA (4,1,2)	1.457	1.857	-0.119	4.377
	SARMA (4,0,2)	1.601	2.001	-0.299	4.806
USA	<b>ARMA (9,0,9)</b>	<b>0.637</b>	<b>0.768</b>	<b>0.412</b>	<b>4.525</b>
	ARIMA (8,1,9)	0.695	0.893	0.205	5.057
	SARIMA (1,1,1)	0.685	0.826	0.321	4.887
	SARMA (3,0,7)	0.657	0.806	0.353	4.618
EU27 & UK	<b>ARMA (7,0,6)</b>	<b>0.446</b>	<b>0.537</b>	<b>0.719</b>	<b>6.186</b>
	ARIMA (7,1,7)	0.620	0.758	0.439	8.072
	SARIMA (5,1,4)	0.546	0.650	0.586	7.386
	SARMA (5,0,6)	0.492	0.601	0.647	6.747
India	<b>ARMA (8,0,9)</b>	<b>0.506</b>	<b>0.602</b>	<b>-0.684</b>	<b>6.287</b>
	ARIMA (7,1,9)	0.622	0.730	-1.469	7.723
	SARIMA (2,1,3)	0.557	0.667	-1.064	6.917
	SARMA (2,0,2)	0.625	0.736	-1.506	7.763

The red mark indicates the best models



**Fig. 7** Plots of actual and predicted emissions for (a) China, (b) USA, (c) EU27&UK, and (d) India using the best statistical model

spurious relationships and yielding inaccurate predictions (Hyndman 2018). By applying differencing to stabilise the time series mean, the models can better assume a constant mean, thereby improving their predictive accuracy.

Seasonal patterns in daily CO<sub>2</sub> emissions significantly influence regional variations in model performance. In regions such as China and the USA, emissions peak during December and January due to increased energy consumption for heating during the winter months. These seasonal fluctuations are more pronounced in China because of its larger industrial base and reliance on coal-powered energy sources (Zhang et al. 2019). Conversely, the EU27&UK region also displays clear seasonal peaks, although these tend to be less pronounced than those in China. This can be attributed to the more diversified energy sources and stringent emission regulations within the EU, which have helped curb emissions (European Environment Agency 2021). In contrast, India shows a comparatively less pronounced seasonal pattern, which may be due to lower variability in energy demand throughout the year, owing to its tropical climate and lower per capita energy consumption (Vishwanathan and Garg 2020).

The seasonal patterns of daily CO<sub>2</sub> emissions introduce significant nonlinearity and non-stationarity into the datasets, complicating the prediction task for the machine learning models. The Augmented Dickey-Fuller (ADF) test confirmed the initial non-stationarity of the USA, the EU27&UK, and India datasets, which required first-order differencing to achieve stationarity. By stabilising the time series through differencing, the models were able to better handle these seasonal effects, leading to improved predictive performance. Moreover, ML models often benefit from

feature engineering (Fan et al. 2019; Wang et al. 2022). While differencing may sometimes result in the loss of valuable information, it also creates new complex features and patterns that ML models can capture, thereby enhancing the model's performance.

Additionally, the model's performance improved when ensemble techniques such as bagging and voting were applied, increasing the  $R^2$  value from a range of 0.677–0.906 to 0.720–0.926 across the four regions. Overall, the ML model performed well in predicting daily emissions, as demonstrated by the fitting curve of the best ML model across each region, as illustrated in Fig. 8. The best-performing ML models are the bagging model (China/India dataset) and the voting model (USA/EU27&UK dataset). These results highlight the importance of handling seasonality and non-stationarity in time series forecasting to improve the accuracy of CO<sub>2</sub> emission predictions.

### Deep learning models

Since differencing improves the performance of ML models, DL models (ANN LSTM, BiLSTM, GRU, and hybrid models) were also applied to the differenced dataset. Notably, the hybrid combination of CNN-RNN slightly improved the performance of the RNN models for all regions except for the Indian dataset, where the performance remained relatively unchanged. As discussed in the “Dataset” section, India's dataset exhibits the least variability and seasonality compared with other regions, as indicated by its lower standard deviation. This reduced variability limits the complexity that DL models, such as CNN-RNN hybrids, can leverage to capture intricate patterns. Additionally, while first-order



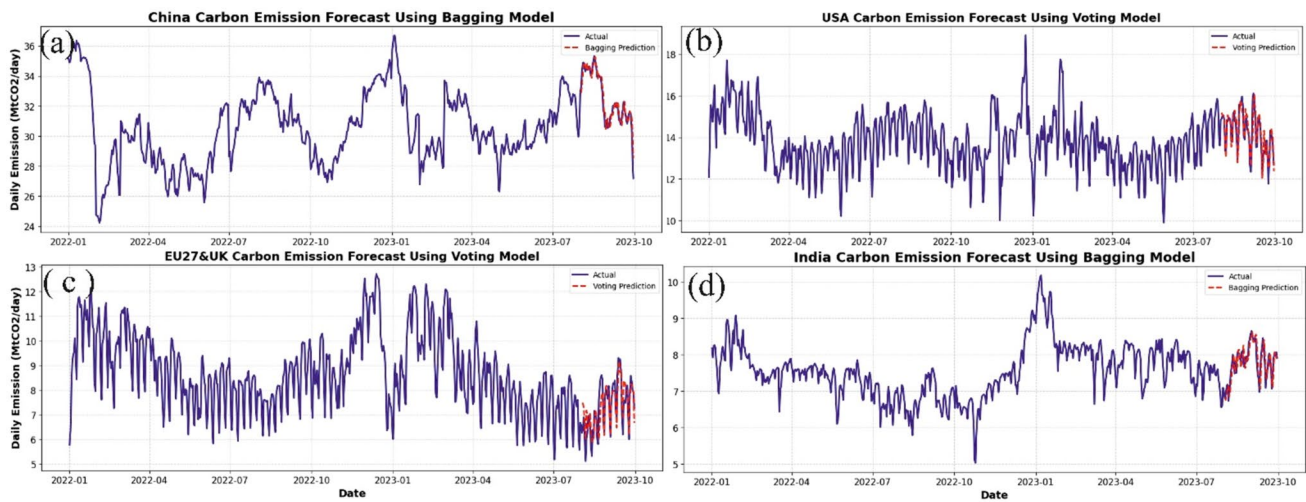
**Table 8** Machine learning models performance evaluation

Region	Model	MAE	RMSE	R <sup>2</sup>	MAPE
China (No differencing)	RF	0.409	0.556	0.899	1.285
	SVR	0.451	0.618	0.876	1.401
	GB	0.419	0.562	0.898	1.314
	Voting	0.399	0.538	0.906	1.252
	Bagging	0.405	0.551	0.901	1.272
China (with differencing)	RF	0.372	0.503	0.918	1.172
	SVR	0.352	0.486	0.923	1.109
	GB	0.374	0.516	0.913	1.180
	Voting	0.351	0.482	0.925	1.108
	Bagging	0.349	0.477	0.926	1.101
USA (No differencing)	RF	0.518	0.645	0.585	3.691
	SVR	0.458	0.559	0.689	3.249
	GB	0.525	0.671	0.551	3.739
	Voting	0.458	0.559	0.688	3.251
	Bagging	0.461	0.562	0.685	3.265
USA (With differencing)	RF	0.323	0.444	0.804	2.279
	SVR	0.358	0.474	0.776	2.517
	GB	0.318	0.458	0.791	2.258
	Voting	0.322	0.442	0.806	2.287
	Bagging	0.325	0.447	0.801	2.297
EU27&UK (No differencing)	RF	0.453	0.604	0.644	6.377
	SVR	0.535	0.689	0.536	7.953
	GB	0.468	0.606	0.641	6.606
	Voting	0.444	0.579	0.672	6.250
	Bagging	0.458	0.605	0.643	6.396
EU27&UK (With differencing)	RF	0.319	0.389	0.852	4.427
	SVR	0.322	0.391	0.850	4.389
	GB	0.323	0.386	0.854	4.479
	Voting	0.314	0.377	0.861	4.322
	Bagging	0.318	0.387	0.854	4.337
India (No differencing)	RF	0.220	0.269	0.663	2.878
	SVR	0.215	0.264	0.678	2.801
	GB	0.219	0.268	0.668	2.869
	Voting	0.218	0.263	0.679	2.842
	Bagging	0.215	0.264	0.677	2.805
India (With differencing)	RF	0.205	0.248	0.714	2.659
	SVR	0.209	0.256	0.696	2.715
	GB	0.209	0.259	0.690	2.720
	Voting	0.202	0.247	0.717	2.627
	Bagging	0.201	0.246	0.720	2.605

The red mark indicates the best models

differencing successfully addresses non-stationarity, the lack of rich seasonal and nonlinear patterns in India's data may have restricted the models from achieving further performance improvements. As detailed in Table 9, the  $R^2$  values of the CNN-GRU model improved from 0.905 to 0.923, 0.755 to 0.803, and 0.841 to 0.878 for China, the USA, and the EU27&UK, respectively. According to Duan et al.

(2022), this blend enables the model to acquire temporal and spatial characteristics from the data, enhancing overall performance. Overall, the predictive accuracy of the DL model is closely aligned with the actual data, as shown in Fig. 9. The best-performing DL models were ANN (for the China and India datasets), CNN-BILSTM (for the USA dataset), and CNN-LSTM (for the EU27&UK datasets).



**Fig. 8** Plots of actual and predicted emissions for (a) China, (b) USA, (c) EU27&UK, and (d) India using the best ML models

**Table 9** Machine learning models performance evaluation

Region	Model	MAE	RMSE	R <sup>2</sup>	MAPE
China	<b>ANN</b>	<b>0.339</b>	<b>0.466</b>	<b>0.930</b>	<b>1.066</b>
	LSTM	0.363	0.497	0.920	1.141
	BILSTM	0.379	0.513	0.915	1.191
	GRU	0.398	0.543	0.905	1.255
	CNN-LSTM	0.385	0.525	0.911	1.212
	CNN-BILSTM	0.347	0.483	0.924	1.092
	CNN-GRU	0.348	0.484	0.923	1.097
USA	ANN	0.355	0.480	0.772	2.506
	LSTM	0.347	0.448	0.799	2.429
	BILSTM	0.389	0.480	0.771	2.755
	GRU	0.391	0.496	0.755	2.758
	CNN-LSTM	0.346	0.447	0.801	2.436
	<b>CNN-BILSTM</b>	<b>0.349</b>	<b>0.443</b>	<b>0.804</b>	<b>2.468</b>
	CNN-GRU	0.352	0.444	0.803	2.494
EU27 & UK	ANN	0.311	0.374	0.863	4.304
	LSTM	0.301	0.359	0.874	4.158
	BILSTM	0.302	0.368	0.868	4.208
	GRU	0.319	0.403	0.841	4.442
	<b>CNN-LSTM</b>	<b>0.289</b>	<b>0.349</b>	<b>0.881</b>	<b>3.999</b>
	CNN-BILSTM	0.303	0.356	0.876	4.253
	CNN-GRU	0.284	0.353	0.878	3.972
India	<b>ANN</b>	<b>0.183</b>	<b>0.235</b>	<b>0.743</b>	<b>2.373</b>
	LSTM	0.206	0.250	0.710	2.685
	BILSTM	0.205	0.248	0.716	2.685
	GRU	0.211	0.252	0.707	2.751
	CNN-LSTM	0.199	0.248	0.715	2.597
	CNN-BILSTM	0.212	2.551	0.696	2.752
	CNN-GRU	0.211	0.256	0.696	2.761

The red mark indicates the best models

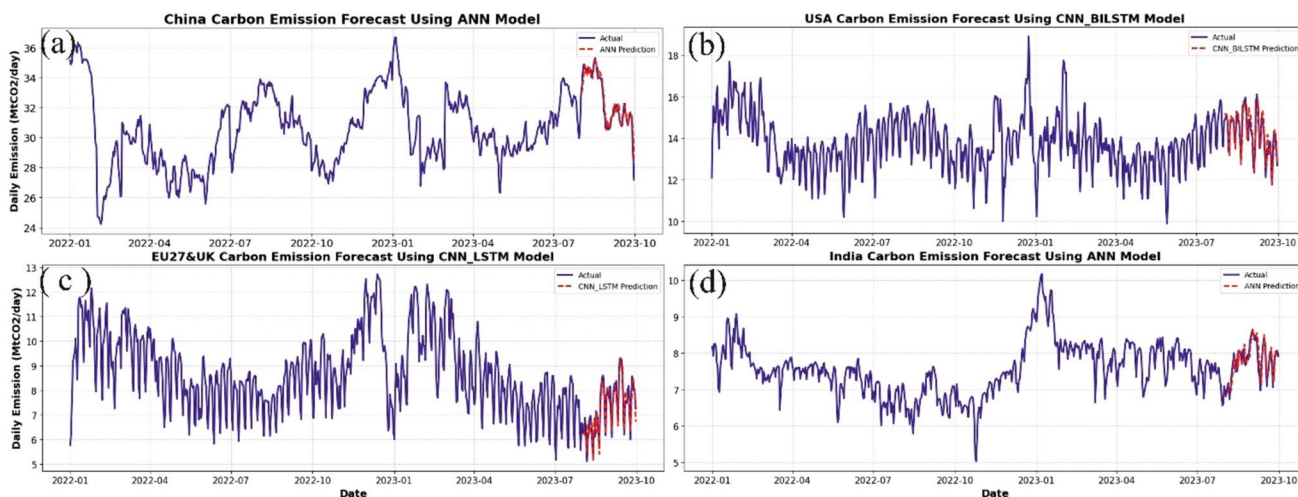


Fig. 9 Plots of actual and predicted emissions for (a) China, (b) USA, (c) EU27&UK, and (d) India using the best DL models

**Discussion**

**Comparative analysis of the performance of the best statistical, ML, and DL models**

Figure 10 compares the best statistical, ML, and DL models across the four regions. The results show that both the ML

and DL models consistently outperform traditional statistical models in predicting daily CO<sub>2</sub> emissions, with lower MAE and RMSE values and higher R<sup>2</sup> scores. Statistical models, with R<sup>2</sup> values ranging from -0.060 to 0.719 and RMSE values between 1.695 and 0.537, struggled to capture the complex, nonlinear patterns present in the emissions data. In contrast, the ML and DL models achieved R<sup>2</sup> values from

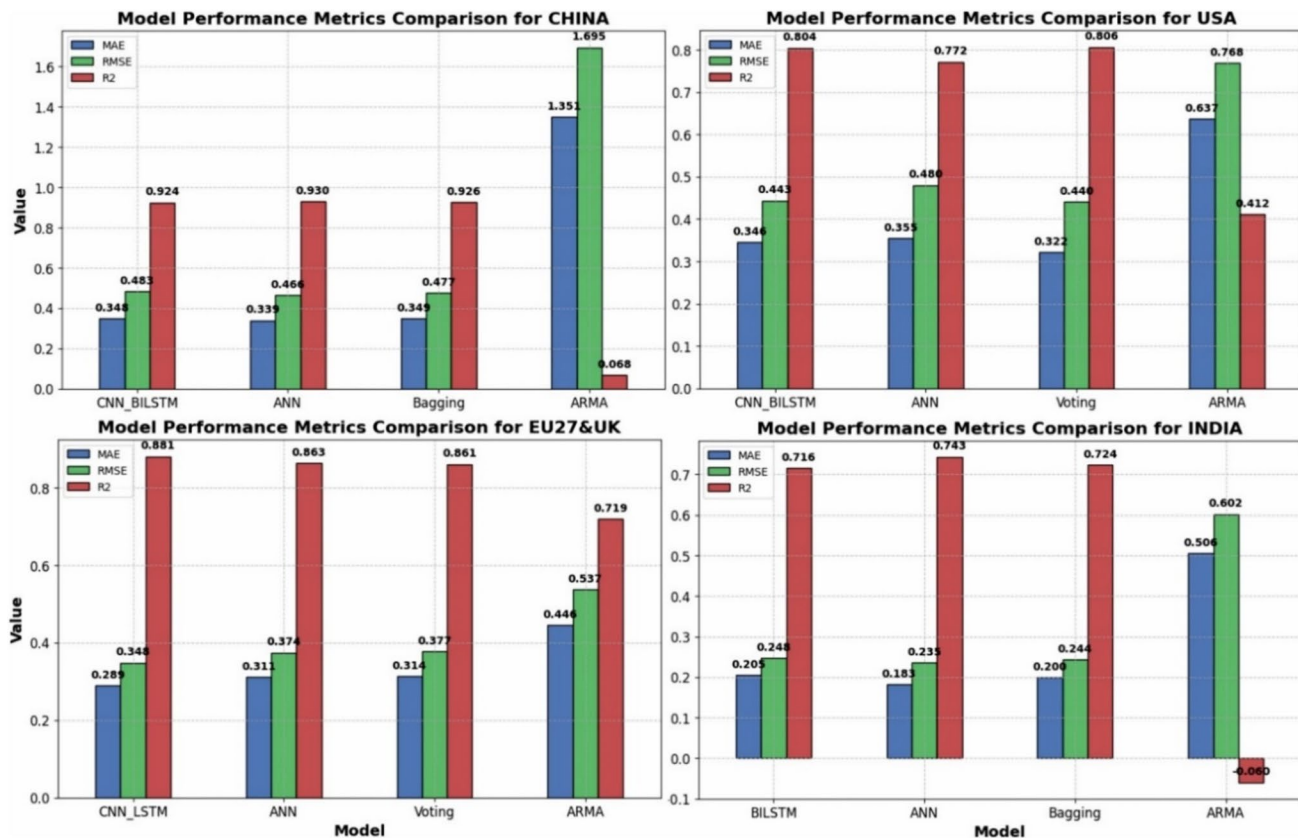


Fig. 10 Comparative assessment of the best statistical, ML, and DL models across the four regions

0.714 to 0.932 and RMSE values as low as 0.247, indicating more accurate and robust predictions. These findings align with the research of Li and Zhang (2023), who reported that ML and DL models are better equipped to model nonlinear datasets, as is the case with daily CO<sub>2</sub> emissions. This ability to handle nonlinearities likely explains the superior performance of the ML and DL models in our study.

The statistical models struggled to capture the high variability and nonlinearity inherent in the emission datasets. These models rely on assumptions of stationarity, which are often violated in real-world CO<sub>2</sub> emissions data due to fluctuating economic activities, weather conditions, and policy changes. The inability of statistical models to adequately handle these dynamic and nonlinear factors explains their lower performance in this study. Future work could explore hybrid approaches that combine statistical models with ML/DL techniques to manage these complexities better, potentially using statistical models for short-term forecasts and ML/DL models for long-term predictions. Such an approach would leverage the strengths of both methods, offering more flexibility in handling different data characteristics.

Model performance varies across regions, suggesting that the specific characteristics of each dataset influence the best-performing model. For example, the ANN model performed best for the China and India datasets, while the ML-voting model performed excellently for the USA dataset, and the CNN-LSTM model achieved the highest accuracy for the EU27&UK dataset. These variations can be attributed to differences in regional data characteristics, such as the degree of nonlinearity, variability, and seasonality. As noted by Oreski et al. (2017), model performance is highly dependent on the data attributes, reinforcing the need for region-specific modelling approaches.

Furthermore, the performances of the ML and DL models were relatively similar. For example, metrics for CNN-BILSTM (RMSE = 0.483,  $R^2 = 0.924$ ), ANN (RMSE = 0.466,  $R^2 = 0.930$ ), and ML-bagging (RMSE = 0.477,  $R^2 = 0.926$ ) were reported for the China dataset. This similarity may be attributed to the relatively small sample size (638 data points) used in this study. When trained on larger datasets, deep learning models typically show a clear advantage over ML models, as they can learn more complex patterns over time. Romeiko et al. (2020) reported that the performance gap between ML and DL models widens with increasing dataset size, suggesting that DL models would likely outperform ML models with larger datasets.

The computational complexity of deep learning (DL) models, which often require specialised hardware such as GPUs or TPUs, may limit their practical application in environments with limited resources. Following the principle of **Occam's Razor**, which suggests favouring simpler

solutions when they achieve comparable results, machine learning (ML) models provide a compelling alternative. ML models, particularly ensemble techniques such as voting and bagging, offer a more computationally efficient approach while maintaining high predictive accuracy. This balance between performance and computational demand makes ML models more suitable for real-world applications where data or hardware resources are limited. As shown in Fig. 11, the ML models were used to forecast daily emissions for the next 60 days across the four regions. The forecast indicates a similar trend to that of the previous winter period, with a gradual increase in CO<sub>2</sub> emissions as regions progress deeper into winter. This increase in emissions is attributed to the heightened demand for heating, which leads to greater coal consumption and elevated CO<sub>2</sub> emissions. These insights are crucial for policymakers in anticipating seasonal spikes in emissions and implementing appropriate mitigation strategies.

### Comparison with previous studies

Compared with our study, Li and Zhang (2023) reported an  $R^2$  of 0.96 when LSTM was used to forecast daily emissions from China during the COVID-19 period, requiring 3000 epochs for training. In contrast, our study achieved an  $R^2$  of 0.93 with only 200 epochs, demonstrating the computational efficiency of our approach. Additionally, our use of differencing significantly improved model accuracy, a factor not considered in previous studies. This enhancement reduces computational demand while maintaining high predictive accuracy, making it suitable for real-time applications in policymaking.

Other studies have also applied machine learning and deep learning techniques to emission prediction, often with a narrower geographic focus or different timeframes. For example, Huang et al. (2019) focused on carbon emissions in China, whereas Song et al. (2023) examined emissions in South Korea. These studies used models on an annual or regional scale. In contrast, our study provides a broader and more dynamic assessment of daily emissions data across four major polluting regions: China, India, the USA, and the EU27&UK. This geographic breadth allows for a more comprehensive analysis, enabling policymakers to set both short- and long-term emission reduction targets.

Furthermore, the robust performance of our models on smaller datasets makes them suitable for less developed countries where data collection and availability are limited, allowing these regions to better participate in global climate change mitigation efforts. This scalability and adaptability set our study apart from others that focused on larger, more comprehensive datasets. Finally, unlike previous studies

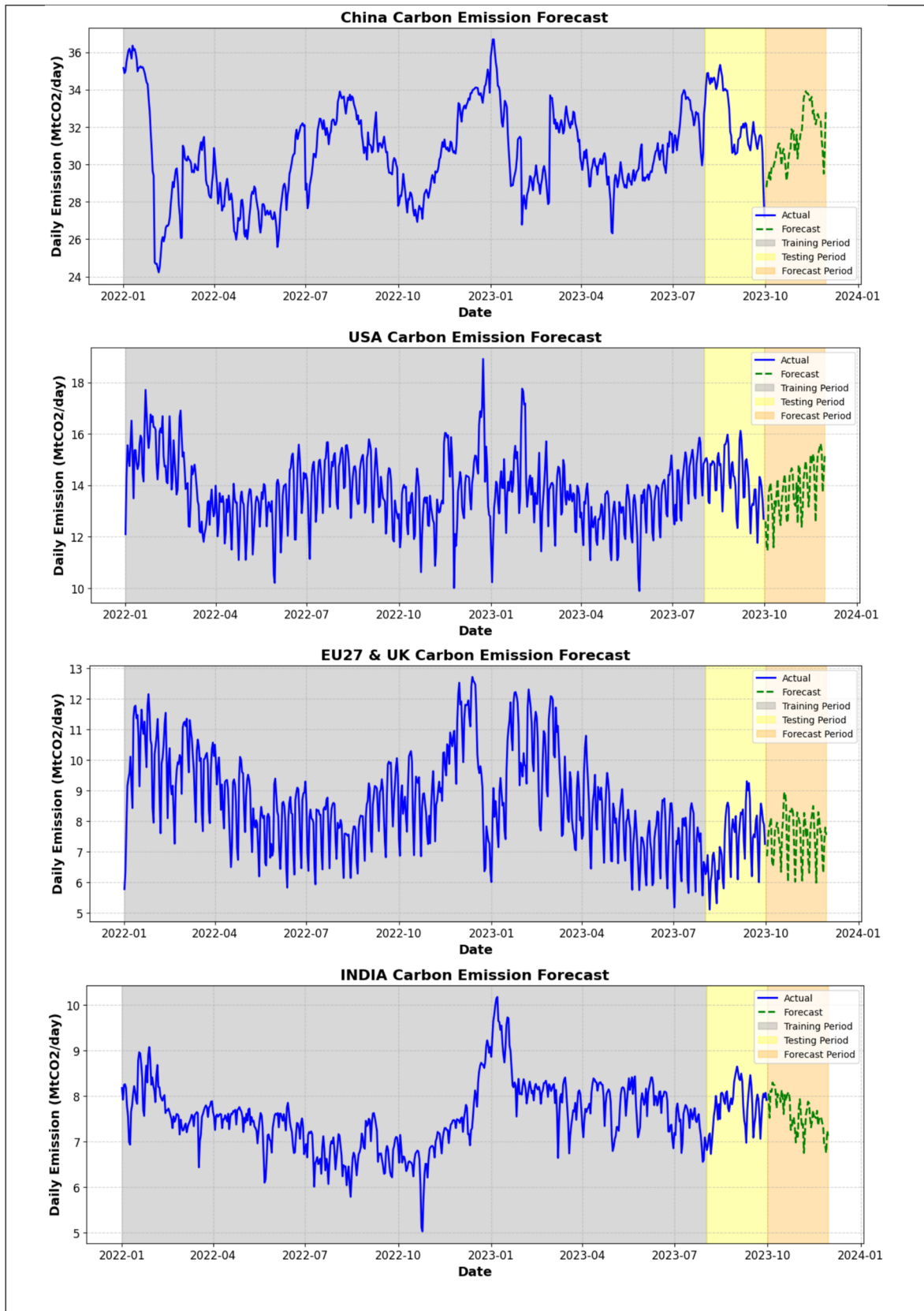


Fig. 11 CO<sub>2</sub> emissions forecast for the next 60 days across the four regions

that focused predominantly on annual predictions, our use of daily emissions data provides timely insights into emissions trends. This real-time capability is essential for governments aiming to make rapid policy adjustments, particularly during periods of economic recovery or energy transition, such as those faced by the major polluting regions analysed in this study.

## Conclusion

The accurate prediction of daily real-time CO<sub>2</sub> emissions holds significant importance for governmental initiatives to mitigate global warming. This study conducted a comparative analysis of the performance of four statistical, three ML, and seven DL models in predicting daily CO<sub>2</sub> emissions across four regions. The models were evaluated using four evaluation metrics, and the following conclusions were drawn from this study.

- The ML and DL models, with higher  $R^2$  (0.714–0.932) and lower RMSE (0.480–0.247) values, respectively, outperformed the statistical model, which had  $R^2$  (–0.060–0.719) and RMSE (1.695–0.537) values, in predicting daily CO<sub>2</sub> emissions across all four regions.
- The performance of the ML models significantly improved through a combination of differencing and the application of ensemble techniques (voting and bagging), resulting in an average increase of 9.6% in  $R^2$ , expanding the range from 0.677–0.906 to 0.720–0.926 across all four regions.
- The performance of the RNN models was enhanced by a hybrid combination of CNN-RNN, which improved the  $R^2$  values across all the regions and increased the range from 0.710–0.923 to 0.743–0.930.

Overall, the performances of both the ML and DL methods were relatively similar. However, owing to the high computational requirements associated with DL models, the recommended models for daily CO<sub>2</sub> emission prediction are ML models that use the ensemble technique of voting and bagging. These models are then employed to forecast the emission trends across the four regions for the next 60 days. This study contributes to the expanding body of research on carbon emission prediction, offering valuable insights into the effectiveness of short-term CO<sub>2</sub> emission forecasts. These findings have significant implications for policymakers, enabling more dynamic adjustments to emission reduction strategies based on daily data and ultimately supporting more effective global warming mitigation efforts.

## Policy implications

Using machine learning and deep learning models to predict daily CO<sub>2</sub> emissions can enable timely policy adjustments and help prevent emission spikes, especially during economic recovery phases. As regions such as China, India, the USA, and the EU strive to meet carbon reduction targets, accurate daily emissions data will equip governments with tools to set short-term targets and make real-time interventions, allowing them to adjust policies more frequently than annual predictions allow.

For example, as economies recover post-COVID-19, monitoring real-time emissions data can prevent emissions from rebounding to pre-crisis levels, ensuring progress toward carbon neutrality. Governments should use these data to implement targeted policies in high-emission sectors such as energy, transportation, and industry:

- Energy sector:** Policies that encourage the expansion of renewable energy sources (such as solar, wind, and biomass) are vital. Promoting the use of clean energy over fossil fuels will help lower emissions in energy-intensive countries such as China and the USA.
- Transportation sector:** Investment in sustainable transport infrastructure, the widespread adoption of electric vehicles, and green public transportation systems significantly contribute to reducing emissions from ground transportation, a major source of CO<sub>2</sub> in regions such as the EU and the USA.
- Industrial sector:** Adopting clean technology and energy-efficient practices in manufacturing and heavy industries can help curb emissions in countries with high industrial output, such as China and India.

Additionally, promoting low-carbon lifestyles and raising public awareness of energy conservation will further support emission reductions. This can be achieved through educational campaigns, financial incentives for energy-saving practices, and the integration of sustainability principles into everyday life. Governments must actively encourage the adoption of energy-efficient practices across all sectors by implementing supportive regulations and offering incentives for innovation. These efforts will align policies with sustainable development goals, ultimately contributing to a greener, more resilient future.

## Limitations and future work

Despite the advancements in daily CO<sub>2</sub> emissions prediction demonstrated in this study, reliance on univariate data is a

key limitation. Important exogenous factors were not considered, such as daily renewable energy consumption, population data, GDP, and daily pump prices, which significantly influence CO<sub>2</sub> emissions. The exclusion of these factors may result in less comprehensive predictions. A major challenge is that many of these variables are only available as monthly or yearly data, making them difficult to incorporate into daily models.

Our future research will address this limitation by employing a multivariate approach, as Li and Zhang (2023) suggested, which would allow for incorporating these exogenous factors, thus potentially enhancing the models' accuracy and robustness. This approach is particularly effective when hybrid machine learning and deep learning techniques are utilised, as Juliet et al. (2024) demonstrated. Furthermore, developing methods to integrate these diverse data types could result in a more precise and comprehensive model for predicting emissions. In addition, future work could explore hybrid approaches that combine statistical models with ML/DL techniques to better manage these complexities, potentially using statistical models for short-term forecasts and ML/DL models for long-term predictions. Such an approach would leverage the strengths of both methods, offering more flexibility in handling different data characteristics.

Additionally, while this study focuses on the overall CO<sub>2</sub> emissions from China, India, the USA, and the EU27&UK, future work will aim to explore the sectoral contributions to daily emissions across these regions. Specifically, we will investigate how different sectors (aviation, ground transport, power, industrial, and residential) contribute to emission patterns and how these contributions vary across regions. Understanding these sectoral contributions will provide detailed insights that are essential for developing region-specific emission reduction strategies.

Moreover, the scope of our future studies will expand to a more global or continental level, exploring daily CO<sub>2</sub> emission trends on a broader scale. This would lay the groundwork for incorporating short-term CO<sub>2</sub> mitigation policies globally, providing critical insights into emission trends across different regions and facilitating the creation of more effective global policies. In summary, while ML and deep learning models have shown promise, future research should move beyond the limitations of univariate models and focus on multivariate approaches to improve accuracy and policy relevance.

## Appendix

**Table 10** Nomenclature list

Symbol	Nomenclature
CO <sub>2</sub>	Carbon Dioxide
MtCO <sub>2</sub> /day	Million Tons of CO <sub>2</sub> per Day
ML	Machine Learning
DL	Deep Learning
ARIMA	Autoregressive Integrated moving average
SARIMA	Seasonal Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
ANN	Artificial Neural Network
SVM	Support Vector Machine
RF	Random Forest
GB	Gradient Boosting
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
BILSTM	Bidirectional Long Short-Term Memory
R <sup>2</sup>	Coefficient of Determination
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ADF	Augmented Dickey-Fuller Test
PACF	Partial Autocorrelation Function
ACF	Autocorrelation Function
$f_t$	Forget Gate
$i_t$	Input Gate
$O_t$	Output Gate
$r_t$	Reset Gate
$z_t$	Update Gate
$C_t$	Memory Cell

**Acknowledgements** The authors thank Dr. O.J. Davis for his constructive discussions and insights during this project. The authors also extend their gratitude to the Centre of Excellence for Data Science, Artificial Intelligence, and Modelling (DAIM) at the University of Hull for providing an academic environment that fostered the development and implementation of the ideas presented in this study.

**Author contribution** Ajala A.A.: Conceptualisation, Result interpretation, and Manuscript draft; Adeoye O.L.: Data collection, Methodology, and Coding; Salami OM: Literature review and Code verification; Jimoh Y.A.: Manuscript correction and editing.

**Data availability** The dataset used for this study is available from the Carbon Monitor project repository (<https://carbonmonitor.org>) and is made freely available to the public under a fair use open data policy.

**Code availability** The code used for the analysis is available upon request to enable full reproducibility of the model implementation and results.

## Declarations

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication** I, Ajala A.A., the corresponding author of this article, hereby confirm that all coauthors have agreed on the final version of this manuscript to be published according to the journal guidelines.

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adegboye OR, Feda, AK, Agyekum EB, Mbasso WF, Kamel S (2024) Towards greener futures: SVR-based CO<sub>2</sub> prediction model boosted by SCMSSA algorithm. *Heliyon*, 10(11). <https://doi.org/10.1016/j.heliyon.2024.e31766>
- Ağbulut Ü (2022) Forecasting of transportation-related energy demand and CO<sub>2</sub> emissions in Turkey with different machine learning algorithms. *Sustain Prod Consum* 29:141–157. <https://doi.org/10.1016/j.spc.2021.10.001>
- Ağbulut Ü, Gürel AE, Biçen Y (2021) Prediction of daily global solar radiation using different machine learning algorithms: evaluation and comparison. *Renew Sustain Energy Rev* 135:110114. <https://doi.org/10.1016/j.rser.2020.110114>
- Ahmed M, Shuai C, Ahmed M (2023) Analysis of energy consumption and greenhouse gas emissions trend in China, India, the USA, and Russia. *Int J Environ Sci Technol* 20(3):2683–2698. <https://doi.org/10.1007/s13762-022-04159-y>
- Ajewole K, Adejuwon S, Jemilohun V (2020) Test for stationarity on inflation rates in Nigeria using augmented Dickey Fuller test and Phillips-persons test. *J Math* 16:11–14
- Amarpuri L, Yadav N, Kumar G, Agrawal S (2019) Prediction of CO<sub>2</sub> emissions using deep learning hybrid approach: a case study in Indian context, 2019 twelfth international conference on contemporary computing (IC3). IEEE. <https://doi.org/10.1109/IC3.2019.8844902>
- Aamir M, Bhatti MA, Bazai SU, Marjan S, Mirza AM, Wahid A, Bhatti UA (2022) Predicting the environmental change of carbon emission patterns in South Asia: a deep learning approach using BiLSTM. *Atmosphere* 13(12):2011. <https://doi.org/10.3390/atmos13122011>
- Azevedo I, Bataille C, Bistline J, Clarke L, Davis S (2021) Net-zero emissions energy systems: what we know and do not know. *Energy and Climate Change* 2:100049. <https://doi.org/10.1016/j.egycc.2021.100049>
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Cho M, Hwang J, Chen C (1995) Customer short-term load forecasting by using ARIMA transfer function model, Proceedings 1995 International Conference on Energy Management and Power Delivery EMPD'95. IEEE. <https://doi.org/10.1109/EMPD.1995.500746>
- Crippa M, Guizzardi D, Solazzo E, Muntean M, Schaaf E, Monforti-Ferrario F, Banja M, Olivier J, Grassi G, Rossi S (2021) GHG emissions of all world countries, EUR 30831 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-41546-6. <https://doi.org/10.2760/173513>
- Ding C, Zhou Y, Pu G, Zhang H (2022) Low carbon economic dispatch of power system at multiple time scales considering GRU wind power forecasting and integrated carbon capture. *Front Energy Res* 10:953883. <https://doi.org/10.3389/fenrg.2022.953883>
- Duan J, Chang M, Chen X, Wang W, Zuo H, Bai Y, Chen B (2022) A combined short-term wind speed forecasting model based on CNN-RNN and linear regression optimisation considering error. *Renew Energy* 200:788–808. <https://doi.org/10.1016/j.renene.2022.09.114>
- European Environment Agency (2021) Trends and Projections in Europe 2021: tracking progress towards Europe's climate and energy targets. EEA Report No 13/2021. <https://doi.org/10.2800/537176>
- Fakana ST (2020) Causes of climate change. *Glob J Sci Front Res* 20:7–12
- Fan F, Lei Y (2017) Index decomposition analysis on factors affecting energy-related carbon dioxide emissions from residential consumption in Beijing. *Math Probl Eng* 2017(1):4963907. <https://doi.org/10.1155/2017/4963907>
- Fan C, Sun Y, Zhao Y, Song M, Wang J (2019) Deep learning-based feature engineering methods for improved building energy prediction. *Appl Energy* 240:35–45. <https://doi.org/10.1016/j.apenergy.2019.02.052>
- Faruque MO, Rabby MAJ, Hossain MA, Islam MR, Rashid MMU, Muyeen SM (2022) A comparative analysis to forecast carbon dioxide emissions. *Energy Rep* 8:8046–8060. <https://doi.org/10.1016/j.egyr.2022.06.025>



- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 1189–1232. <https://www.jstor.org/stable/2699986>
- Geevaretnam JL, Zainuddin NMM, Kamaruddin N, Rusli H, Maarop N, Hassan WAW (2022) Predicting the carbon dioxide emissions using machine learning. *Int J Innov Comput* 12(2):17–23. <https://doi.org/10.11113/ijic.v12n2.369>
- Giannelos S, Moreira A, Papadaskalopoulos D, Borozan S, Pudjianto D, Konstantelos I, Strbac G (2023) A machine learning approach for generating and evaluating forecasts on the environmental impact of the buildings sector. *Energies* 16(6):2915. <https://doi.org/10.3390/en16062915>
- Guo LN, She C, Kong DB, Yan SL, Xu YP, Khayatnezhad M, Gholinia F (2021) Prediction of the effects of climate change on hydroelectric generation, electricity demand, and emissions of greenhouse gases under climatic scenarios and optimised ANN model. *Energy Rep* 7:5431–5445. <https://doi.org/10.1016/j.egy.2021.08.134>
- Haider SA, Sajid M, Sajid H, Uddin E, Ayaz Y (2022) Deep learning and statistical methods for short-and long-term solar irradiance forecasting for Islamabad. *Renew Energy* 198:51–60. <https://doi.org/10.1016/j.renene.2022.07.136>
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Huang Y, Shen L, Liu H (2019) Grey relational analysis, principal component analysis and forecasting of carbon emissions based on long short-term memory in China. *J Clean Prod* 209:415–423. <https://doi.org/10.1016/j.jclepro.2018.10.128>
- Huang Z, Wang J, Bing L, Qiu Y, Guo R, Yu Y, Ma M, Niu L, Tong D, Andrew RM (2023) Global carbon uptake of cement carbonation accounts 1930–2021. *Earth Syst Sci Data* 15(11):4947–4958. <https://doi.org/10.5194/essd-15-4947-2023>
- Hyndman RJ (2018) Forecasting: principles and practice. Otexts.
- Jeniffer SB, Sathya D, Yogeshwari R, Kumar SS, Krishna ES, Pranesh MP (2023) Machine learning assisted experimental investigation of carbon emissions of diverse fuels. In 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA) (pp. 128–132). IEEE.
- Juliet AH, Malathi P, Legapriyadharshini N (2024) Smart forecasting: harnessing machine learning for accurate CO2 emission predictions. In: Proceedings of the 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO). IEEE, pp 1–7. <https://doi.org/10.1109/ICRITO61523.2024.10522152>
- Kleynhans T, Montanaro M, Gerace A, Kanan C (2017) Predicting top-of-atmosphere thermal radiance using MERRA-2 atmospheric data with deep learning. *Remote Sens* 9(11):1133. <https://doi.org/10.3390/rs9111133>
- Kong F, Song J, Yang Z (2022) A daily carbon emission prediction model combining two-stage feature selection and optimised extreme learning machine. *Environ Sci Pollut Res* 29(58):87983–87997. <https://doi.org/10.1007/s11356-022-21277-9>
- Kour M (2023) Modelling and forecasting of carbon-dioxide emissions in South Africa by using ARIMA model. *Int J Environ Sci Technol* 20(10):11267–11274. <https://doi.org/10.1007/s13762-022-04609-7>
- Kumari S, Singh SK (2022) Machine learning-based time series models for effective CO<sub>2</sub> emission prediction in India. *Environ Sci Pollut Res* 30(55):116601–116616. <https://doi.org/10.1007/s11356-022-21723-8>
- Lee G, Han J (2020) Forecasting gas demand for power generation with SARIMAX models. *Korean Manag Sci Rev* 37:67–78. <https://doi.org/10.7737/KMSR.2020.37.4.067>
- Legg S (2021) IPCC, 2021: climate change 2021-the physical science basis. *Interaction* 49(4):44–45
- Lennan M, Morgera E (2022) The Glasgow climate conference (COP26). *Int J Mar Coast Law* 37(1):137–151
- Li X, Zhang X (2023) A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China. *Environ Sci Pollut Res* 30(55):117485–117502. <https://doi.org/10.1007/s11356-023-30428-5>
- Liu Z, Deng Z, Zhu B, Ciais P, Davis SJ, Tan J, Andrew RM, Boucher O, Arous SB, Canadell JG (2022) Global patterns of daily CO<sub>2</sub> emissions reductions in the first year of COVID-19. *Nat Geosci* 15(8):615–620. <https://doi.org/10.1038/s41561-022-00965-8>
- Magazzino C, Mele M (2022) A new machine learning algorithm to explore the CO<sub>2</sub> emissions-energy use-economic growth trilemma. *Ann Oper Res* 1–19. <https://doi.org/10.1007/s10479-022-04787-0>
- Magazzino C, Mele M, Drago C, Kuşkaya S, Pozzi C, Monarca U (2023) The trilemma among CO<sub>2</sub> emissions, energy use, and economic growth in Russia. *Sci Rep* 13(1):10225. <https://doi.org/10.1038/s41598-023-37251-5>
- Mason K, Duggan J, Howley E (2018) Forecasting energy demand, wind generation and carbon dioxide emissions in Ireland using evolutionary neural networks. *Energy* 155:705–720. <https://doi.org/10.1016/j.energy.2018.04.192>
- Mohammadi K, Shamsirband S, Tong CW, Arif M, Petković D, Ch S (2015) A new hybrid support vector machine-wavelet transform approach for estimation of horizontal global solar radiation. *Energy Convers Manage* 92:162–171. <https://doi.org/10.1016/j.enconman.2014.12.050>
- Moroney L (2019) Sequences, time series and prediction coursera. Available online: <https://www.coursera.org/learn/TensorFlow-sequences-time-series-and-prediction> Accessed 1 Oct 2023
- Nyoni T, Bonga WG (2019) Prediction of CO<sub>2</sub> emissions in India using arima models. *DRJ-J Econ Financ* 4(2):01–10
- Oreski D, Oreski S, Klicek B (2017) Effects of dataset characteristics on the performance of feature selection techniques. *Appl Soft Comput* 52:109–119. <https://doi.org/10.1016/j.asoc.2016.12.023>
- Patel J, Shah S, Thakkar P, Kotecha K (2015) Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Syst Appl* 42(1):259–268. <https://doi.org/10.1016/j.eswa.2014.07.040>
- Prakash A, Singh SK (2024) CO<sub>2</sub> emission prediction from coal used in power plants: a machine learning-based approach. *Iran Journal of Computer Science*, 1–17. <https://doi.org/10.1007/s42044-024-00185-w>
- Peng Y, Li Q, Kong W, Qin F, Zhang J, Cichocki A (2020) A joint optimisation framework to semi-supervised RVFL and ELM networks for efficient data classification. *Appl Soft Comput* 97:106756. <https://doi.org/10.1016/j.asoc.2020.106756>
- Qiao W, Lu H, Zhou G, Azimi M, Yang Q, Tian W (2020) A hybrid algorithm for carbon dioxide emissions forecasting based on improved lion swarm optimiser. *J Clean Prod* 244:118612. <https://doi.org/10.1016/j.jclepro.2019.118612>
- Rehman A, Ma H, Ahmad A, Ozturk I, Chishti MZ (2021) How do climatic change, cereal crops and livestock production interact with carbon emissions? Updated evidence from China. *Environ Sci Pollut Res* 28:30702–30713. <https://doi.org/10.1007/s11356-021-12948-0>

- Ren Y, Suganthan PN, Srikanth N, Amaratunga G (2016) Random vector functional link network for short-term electricity load demand forecasting. *Inf Sci* 367:1078–1093. <https://doi.org/10.1016/j.ins.2015.11.039>
- Romeiko XX, Guo Z, Pang Y, Lee EK, Zhang X (2020) Comparing machine learning approaches for predicting spatially explicit life cycle global warming and eutrophication impacts from corn production. *Sustainability* 12(4):1481. <https://doi.org/10.3390/su12041481>
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
- Sen P, Roy M, Pal P (2016) Application of ARIMA for forecasting energy consumption and GHG emission: a case study of an Indian pig iron manufacturing organisation. *Energy* 116:1031–1038. <https://doi.org/10.1016/j.energy.2016.10.068>
- Shabani E, Hayati B, Pishbahar E, Ghorbani MA, Ghahremanzadeh M (2021) A novel approach to predict CO<sub>2</sub> emission in the agriculture sector of Iran based on Inclusive Multiple Model. *J Clean Prod* 279:123708. <https://doi.org/10.1016/j.jclepro.2020.123708>
- Song C, Wang T, Chen X, Shao Q, Zhang X (2023) Ensemble framework for daily carbon dioxide emissions forecasting based on the signal decomposition–reconstruction model. *Appl Energy* 345:121330. <https://doi.org/10.1016/j.apenergy.2023.121330>
- Tang QY, Zhang CX (2013) Data Processing System (DPS) software with experimental design, statistical analysis, and data mining developed for use in entomological research. *Insect Sci* 20(2):254–260. <https://doi.org/10.1111/j.1744-7917.2012.01519.x>
- Tümer AE, Akkuş A (2018) Forecasting gross domestic product per capita using artificial neural networks with non-economical parameters. *Physica A* 512:468–473. <https://doi.org/10.1016/j.physa.2018.08.047>
- Vapnik V (1998) The support vector method of function estimation. In *Nonlinear modelling: advance black-box technique* (pp. 55–85). [https://doi.org/10.1007/978-1-4615-5703-6\\_3](https://doi.org/10.1007/978-1-4615-5703-6_3)
- Vishwanathan SS, Garg A (2020) Energy system transformation to meet NDC, 2 C, and well below 2 C targets for India. *Clim Change* 162(4):1877–1891. <https://doi.org/10.1007/s10584-019-02616-1>
- Wang C, Li M, Yan J (2023) Forecasting carbon dioxide emissions: application of a novel two-stage procedure based on machine learning models. *J Water Climate Change* 14(2):477–493. <https://doi.org/10.2166/wcc.2023.331>
- Wang Z, Xia L, Yuan H, Srinivasan RS, Song X (2022) Principles, research status, and prospects of feature engineering for data-driven building energy prediction: a comprehensive review. *J Build Eng* 58:105028. <https://doi.org/10.1016/j.jobe.2022.105028>
- Wang J, Yang F, Chen K (2019) Regional carbon emission evolution mechanism and its prediction approach: a case study of Hebei, China. *Environ Sci Pollut Res* 26:28884–28897. <https://doi.org/10.1007/s11356-019-06021-0>
- Yoon J (2021) Forecasting of real GDP growth using machine learning models: gradient boosting and random forest approach. *Comput Econ* 57(1):247–265. <https://doi.org/10.1007/s10614-020-10054-w>
- Zhang K, Huo X, Shao K (2023) Temperature time series prediction model based on time series decomposition and bilstm network. *Mathematics* 11(9):2060. <https://doi.org/10.3390/math11092060>
- Zhang W, Khan A, Huyan J, Zhong J, Peng T, Cheng H (2021) Predicting marshall parameters of flexible pavement using support vector machine and genetic programming. *Constr Build Mater* 306:124924. <https://doi.org/10.1016/j.conbuildmat.2021.124924>
- Zhang H, Peng J, Wang R, Zhang M, Gao C, Yu Y (2023b) Use of random forest based on the effects of urban governance elements to forecast CO<sub>2</sub> emissions in Chinese cities. *Heliyon*, 9(6). <https://doi.org/10.1016/j.heliyon.2023.e16693>
- Zhang Q, Zheng Y, Tong D, Shao M, Wang S, Zhang Y, Hao J (2019) Drivers of improved PM<sub>2.5</sub> air quality in China from 2013 to 2017. *Proc Natl Acad Sci* 116(49):24463–24469. <https://doi.org/10.1073/pnas.1907956116>
- Zhou W, Zeng B, Wang J, Luo X, Liu X (2021) Forecasting Chinese carbon emissions using a novel grey rolling prediction model. *Chaos, Solitons Fractals* 147:110968. <https://doi.org/10.1016/j.chaos.2021.110968>
- Ziel F, Weron R (2018) Day-ahead electricity price forecasting with high-dimensional structures: univariate vs. multivariate modelling frameworks. *Energy Econ* 70:396–420. <https://doi.org/10.1016/j.eneco.2017.12.016>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.