Deep Learning and XAI for Carbon Dioxide Emissions Prediction: Integrating MLP with SHAP and Multi-Policy Scenario Analysis

Ruimin Ma, Qiong Li and Alexander Kovshov*

Faculty of Applied Mathematics and Control Processes, Saint Petersburg State University, Saint Petersburg, 199034, Russia

Revista Internacional
Métodos numéricos
para cálculo y diseño en ingeniería
RIMNI
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
In cooperation with CIMNE

INFORMATION

Keywords:
Carbon dioxide emission prediction
deep learning
multilayer perceptron
neural network
SHapley Additive exPlanations
scenario analysis

DOI: 10.23967/j.rimni.2025.10.76077

SCIPEDIA
Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería
Revista Internacional Métodos numéricos para cálculo y diseño en ingeniería RIMNI

Deep Learning and XAI for Carbon Dioxide Emissions Prediction: Integrating MLP with SHAP and Multi-Policy Scenario Analysis

Ruimin Ma, Qiong Li and Alexander Kovshov[*]

Faculty of Applied Mathematics and Control Processes, Saint Petersburg State University, Saint Petersburg, 199034, Russia

ABSTRACT

This study is conducted in response to the increasingly prominent climate crisis in contemporary society. It aims to contribute to the growing body of research on the application of deep learning (DL) in environmental sciences and to provide practical guidance for model selection in similar predictive tasks. To this end, the study focuses on carbon dioxide (CO_2) emissions prediction, employing Multilayer Perceptron (MLP) models to analyze multi-country panel data. By integrating MLP with explainable artificial intelligence (XAI) techniques, this research not only investigates the underlying mechanisms of various factors influencing CO_2 emissions but also quantifies and visualizes the contribution of different driving factors to the prediction outcomes, providing decision support for climate governance strategies. Through an analysis of global panel data, we construct a model incorporating 14 driving factors spanning multiple dimensions, including economic, social, environmental, energy, and technology aspects. To optimize the MLP model, we employ a five-dimensional hyperparameter space comprising hidden layer structure, learning rate, batch size, dropout rate, and training epochs and apply Grid Search for parameter tuning. Experimental results indicate that the MLP model achieves R^2 of 0.9951, demonstrating its strong capability in high-precision nonlinear fitting under complex policy scenarios. To further enhance the interpretability of neural networks in CO_2 emissions prediction, we introduce SHapley Additive exPlanations (SHAP) to quantify the marginal contributions of different driving factors. This analysis reveals that energy-related features play a dominant role in emission predictions, laying the foundation for scenario analysis and emission reduction policy evaluation. Furthermore, this study incorporates scenario analysis to simulate potential trajectories of CO_2 emissions under different policy scenarios, providing a quantitative reference for future emission reduction strategies and environmental governance policies.

OPEN ACCESS

Received: 13/11/2025

Accepted: 02/12/2025

Published: 20/03/2026

DOI
10.23967/j.rimni.2025.10.76077

Keywords:
Carbon dioxide emission prediction
deep learning
multilayer perceptron
neural network
SHapley Additive exPlanations
scenario analysis

1 Introduction

Atmospheric concentrations of carbon dioxide (CO_2) continue to rise, not only driving a gradual increase in global temperatures but also exerting profound impacts on both environmental and

[*]Correspondence: Alexander Kovshov (a.kovshov@spbu.ru).

economic systems. This has become one of the most pressing environmental challenges facing humanity [1–3]. In particular, the sustained growth of CO_2 emissions has had far-reaching effects on global warming and environmental degradation [4–7]. Governments and international organizations have taken a range of actions to reduce greenhouse-gas emissions and promote sustainable development [8,9]. For example, the Paris Agreement formalized global temperature-control targets [10] and prompted many countries to establish carbon-neutrality strategies. However, accurately predicting CO_2 emissions remains a crucial yet unresolved issue for supporting policy-making, environmental management, and sustainable development strategies [2,11].

Traditional CO_2 emission prediction methods, such as multiple regression analysis, typically assume linear relationships between variables, which limits their ability to capture the complex nonlinear interactions among factors. Furthermore, CO_2 emissions are influenced by multiple determinants, including economic growth, energy consumption, social structures, and policy environments. Identifying and quantifying the contributions of these factors remains a significant challenge [12–15]. To better understand and respond to the complex impacts of CO_2 emissions, classical machine-learning approaches have shown great potential for modeling complex systems in environmental science and climate-change prediction, and they have achieved some notable results [16–19]. However, limitations persist in feature extraction, utilization of structural information, generalization capacity, and the handling of large-scale data, making it difficult to capture the complex nonlinear coupling effects among driving factors. In addition, much of the existing work focuses predominantly on predictive accuracy while neglecting model interpretability, which hinders decision-makers from understanding the mechanisms of key drivers and thus limits the effectiveness of mitigation strategies [20–23].

In recent years, deep learning (DL) techniques have demonstrated remarkable capabilities in nonlinear modeling across various domains [24–28]. Given that DL can learn high-dimensional feature representations through nonlinear transformations and has not yet been fully explored in CO_2 emission prediction [29], this study employs the Multilayer Perceptron (MLP) model, a classic feedforward neural network that has shown strong performance in complex predictive tasks [30,31]. However, DL models are often considered “black boxes” due to their lack of transparency in decision-making logic, which is particularly concerning for CO_2 emission prediction, where policy sensitivity is high. To address this issue, Explainable Artificial Intelligence (XAI) techniques, such as SHapley Additive exPlanations (SHAP), which is a game-theory-based method, serve as crucial tools for enhancing model interpretability [32–35]. By quantifying the marginal contributions of features to prediction outcomes, SHAP not only reveals the global importance of drivers but also captures their dynamic effects under different scenarios, providing policymakers with a more granular basis for decision-making [36].

Anchored in the increasingly urgent climate-crisis context and targeting the key challenges in current CO_2 forecasting research, this study proposes an innovative approach. While DL exhibits strong nonlinear modeling capabilities, its “black-box” nature constrains transparency and interpretability, limiting its usefulness for policymaking. Furthermore, most studies emphasize static feature analysis and lack scenario-based simulations of dynamic policy interventions, making predictions less adaptable to real-world policy adjustments and energy-structure transitions. At the same time, existing work often falls short on regional heterogeneity analysis, particularly with respect to the deeper mining and application of cross-country panel data.

To address these challenges, we propose a “DL + XAI” paradigm that combines an MLP forecaster with SHAP-based interpretability, built on a dataset covering 107 countries over 20 years [37]. The goal is to jointly optimize forecasting accuracy and transparency. The main contributions are as follows: (1) We model multi-country panel data using an MLP and conduct hyperparameter optimization-covering hidden-layer configurations, learning rate, batch size, dropout rate, and training epochs-to ensure high-accuracy prediction under complex policy environments. (2) We employ SHAP to quantify and visualize how key drivers (economic, social, environmental, energy, and technology) contribute to CO_2 predictions, thereby uncovering their influence mechanisms and providing actionable evidence for policymakers. (3) We design policy-intervention scenarios (e.g., reducing fossil-fuel dependence and increasing the share of renewables) to simulate dynamic changes in CO_2 trajectories and to assess potential impacts, supporting low-carbon transition strategies.

Through this study, we deliver a deep-learning framework for CO_2 emission forecasting that is both accurate and interpretable, thereby laying a data-driven foundation for formulating more scientific and feasible climate-governance policies.

Beyond improving the transparency of the deep model, the SHAP analysis in this study also plays a central role in interpreting the subsequent multi-policy scenario analysis. The proposed "DL + XAI" paradigm operates as an integrated pipeline: a high accuracy MLP is first trained to learn the nonlinear mapping from macroeconomic and energy-related drivers to national CO_2 emissions; SHAP values are then computed to quantify the contribution of each input to the model outputs. When we later construct policy scenarios by perturbing selected, policy-relevant variables, SHAP serves as a principled tool for explaining how these perturbations reweight the contributions of individual drivers and for validating that the simulated trajectories are consistent with the feature-attribution structure learned by the model.

The remainder of this paper is organized as follows. Section 2 reviews related research on CO_2 emission forecasting and the applications of deep learning in this field. Section 3 introduces the research methodology, including the adopted deep-learning model and the feature-importance analysis technique. Section 4 describes the experimental setup, covering data preprocessing, evaluation metrics, hyperparameter settings, and the simulation design for different policy scenarios. Section 5 presents and discusses the results, highlighting their implications. Finally, Section 6 summarizes the main findings and outlines directions for future research.

2 Related Works

The high-precision prediction of CO_2 emissions has become a core issue in environmental science and sustainable development research due to its critical role in global climate governance. Traditional statistical models, such as multiple regression analysis, time series analysis, and econometric models, have been widely applied in CO_2 emissions forecasting. For instance, Sharma [11] conducted a panel data analysis of 69 countries to explore the correlation between economic indicators and CO_2 emissions, while Narayan and Narayan [1] examined the relationship between CO_2 emissions and economic growth in developing countries, establishing a regression model based on economic expansion and energy consumption. Although these models provide a foundational understanding of emissions trends, they rely on the assumption of linear relationships among variables, making them inadequate for capturing the nonlinear coupling effects inherent in energy transitions and technological innovations.

With advancements in ML, ML algorithms have been increasingly adopted for CO_2 emissions forecasting, particularly for their ability to model complex systems. For example, Qin and Gong [5] employed ML techniques to estimate China's CO_2 emissions and identify key driving factors. Wu et al. [38] compared the performance of support vector machines (SVM), random forests, and DL models in CO_2 emissions prediction, concluding that DL methods outperform traditional

ML approaches in capturing nonlinear relationships. However, despite these advancements, existing research still faces several key limitations: (1) Many studies focus on a single country or region, overlooking the impact of global heterogeneity on prediction accuracy. For instance, AlShafeey and Rashdan [23] highlighted that differences in energy structures lead to significant variations in model performance across economies, necessitating the use of cross-country panel data to improve generalizability. (2) While neural networks (e.g., LSTM, MLP) exhibit strong nonlinear modeling capabilities, their "black-box" nature limits their applicability in policy-making. Rolnick et al. [21] emphasized the need to integrate AI techniques with deep learning to enhance interpretability and identify key drivers in complex models. (3) Most existing studies rely on static features, making it difficult to simulate the dynamic pathways of CO_2 emissions under different policy interventions. Rolnick et al. [21] systematically reviewed the applications of ML in climate modeling and policy assessment, highlighting the importance of incorporating dynamic policy responses and multivariate coupling analysis.

In response to these research challenges, the application of deep learning in environmental sciences has seen significant advancements. Lundberg and Lee [39] proposed the SHAP method, which leverages game theory to quantify feature contributions, and has demonstrated success in fields such as healthcare and transportation. However, its application in CO_2 emissions prediction remains in an exploratory stage. This study proposes a synergistic CO_2 emissions prediction framework that integrates MLP with XAI. By employing SHAP, we quantify the contributions of various factors to CO_2 emissions forecasts and develop a multi-policy scenario analysis framework to simulate potential trajectories of CO_2 emissions under different policy interventions.

The novelty of this study lies in the integration of DL's high-precision predictive capabilities with explainability analysis while incorporating policy scenario modeling. This approach addresses existing gaps in model transparency and policy evaluation, thereby advancing the application of DL in environmental sciences and providing data-driven decision support for global climate governance. Compared with our previous work based on tree-based machine learning models, and with existing studies that apply deep neural networks, SHAP or scenario analysis separately, the present paper proposes an integrated "DL + XAI" workflow for multi-country CO_2 emissions. In this framework, an MLP forecaster is calibrated on cross-country panel data, SHAP is then used on the same trained model to decompose the contributions of key drivers, and these XAI results directly guide the construction of multi-policy scenarios, thereby linking numerical prediction, interpretability and policy-oriented simulation in a coherent way.

3 Methodology

For this research area, our previous work, based on tree-based machine learning methods, systematically revealed the global impact of core variables such as economic, social, environmental, energy, and technology on CO_2 emissions. However, tree-based machine learning methods can capture some non-linear relationships, their generalization performance in high-dimensional feature spaces remains constrained.

To overcome these limitations, this study innovatively introduces the MLP model and integrates XAI techniques, specifically SHAP values. Compared to previous models, MLP, with its non-linear transformations through hidden layers and distributed representation learning, significantly improves the modeling accuracy of multi-variable coupling relationships in complex policy scenarios. This is particularly evident when capturing the synergistic effects between features. At the same time, by leveraging SHAP values, a game-theory-driven explanation tool, we can not only quantify the global contribution of each feature to the prediction of CO_2 emissions but also visually present the specific impact of key variables on emissions under different conditions through the decomposition of conditional expectations.

The innovation of this study lies in the combined application of the existing MLP model and SHAP values for CO_2 emission analysis, successfully achieving a collaborative optimization of prediction accuracy and interpretability. The high-dimensional non-linear fitting ability of the MLP model ensures the scientific validity of policy scenario simulations, while the multi-layered interpretative framework of SHAP provides decision-makers with a transparent interpretation of the "black box" model. This "DL + XAI" integration paradigm not only deepens the theoretical understanding of the multi-scale driving mechanisms behind CO_2 emissions but also offers actionable decision support tools for the design of differentiated and dynamic climate governance strategies by establishing quantifiable relationships between policy parameters and emission reduction outcomes.

From a numerical perspective, the proposed methodology can be viewed as an integrated three-step procedure on the multi-country panel dataset $D = \{(X_i, C_i)\}_{i=1}^{N}$, where $X_i \in \mathbb{R}^p$ denotes the input vector of features and $C_i \in \mathbb{R}$ the corresponding CO_2 emissions. First, an MLP $\Phi_\theta : \mathbb{R}^p \to \mathbb{R}$ is trained by solving the empirical risk minimization problem

$$\theta^* = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \left(C_i - \Phi_\theta(X_i)\right)^2, \tag{1}$$

where N is the number of training samples, X_i is the 70-dimensional input feature vector constructed from the five-year window, C_i is the corresponding observed CO_2 emission, and $\Phi_\theta(X_i)$ is the model prediction. The parameter vector θ collects all weights and biases of the MLP and is learned by backpropagation with gradient-based optimization.

3.1 MLP

MLP is a classic feedforward artificial neural network model, consisting of an input layer, hidden layers, and an output layer. MLP enables the network to represent complex function mappings through non-linear activation functions, extracting higher-order features from input data and ultimately modeling complex relationships. It is a commonly used network structure in DL. In this study, the choice of an MLP architecture over more complex deep models such as LSTMs or GRUs is motivated by both data characteristics and numerical considerations. As detailed in Section 4.1, each training sample is constructed as a five-year sliding window of 14 driving factors, which is then flattened into a fixed-length feature vector of dimension 5 × 14 = 70. This representation is essentially tabular, and the sequence length is relatively short, so recurrent architectures that are designed for long sequences do not provide a clear structural advantage, while they would introduce substantially more trainable parameters and a higher risk of overfitting given the moderate sample size ($N = 2247$). The network structure of MLP is shown in Fig. 1.

MLP consists of multiple layers, each made up of several neurons (also called "nodes," such as x_1 in Fig. 1), and the layers are fully connected through weight matrices.

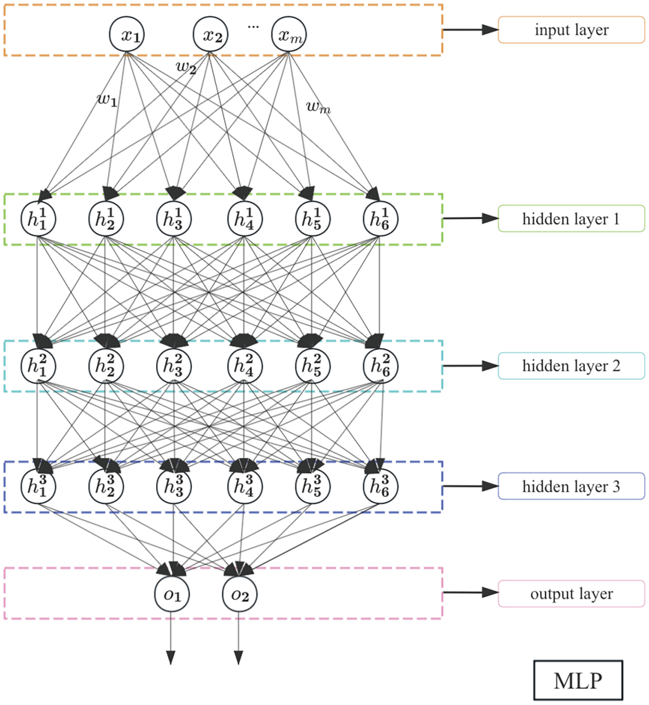
x_1
x_2
...
x_m
w_1
w_2
w_m
input layer
h_1^1 h_2^1 h_3^1 h_4^1 h_5^1 h_6^1
hidden layer 1
h_1^2 h_2^2 h_3^2 h_4^2 h_5^2 h_6^2
hidden layer 2
h_1^3 h_2^3 h_3^3 h_4^3 h_5^3 h_6^3
hidden layer 3
o_1
o_2
output layer
MLP

Figure 1: The network structure of MLP

3.1.1 Input Layer

Contains an m-dimensional feature vector $[x_1, x_2, \ldots, x_m]$ (with m input nodes), corresponding to the carbon emission driving factors in the study. Each input value x_i is connected to its corresponding weight w_i.

3.1.2 Hidden Layers

Composed of several neurons, each of which receives the output from the previous layer's neurons, performs a weighted sum, and then passes the result through an activation function to generate an output. The number of layers and neurons per layer can be adjusted based on the task's complexity. This study uses a classic MLP with three hidden layers. Fig. 1 shows that each hidden layer has 6 neurons, and this setup is used for understanding the principles of MLP.

- Hidden Layer 1: Each neuron is connected to all the nodes in the input layer, performs a weighted sum, and then passes the result through an activation function to obtain the output.
- Hidden Layer 2: Each neuron is connected to all the nodes in hidden Layer 1, performs a weighted sum, and then passes the result through an activation function to obtain the output.
- Hidden Layer 3: Each neuron is connected to all the nodes in hidden Layer 2, performs a weighted sum, and then passes the result through an activation function to obtain the output.

For the j-th neuron in the l-th layer, its calculation formula is given by Eq. (2):

$$h_l^j = f\left(\sum_{i=1}^{m} w_{ij}^{(l)} h_{l-1}^i + b_l^j\right) \tag{2}$$

where h_l^j is the output of the j-th neuron in the l-th layer, h_{l-1}^i is the output of the i-th neuron in the $(l-1)$-th layer, m denotes the number of neurons in the previous layer (i.e., the number of input connections to neuron j), $w_{ij}^{(l)}$ is the weight from neuron i in layer $l-1$ to neuron j in layer l, b_l^j is the bias term, and $f(\cdot)$ is the activation function.

3.1.3 Output Layer

The output layer receives the activations from the last hidden layer and produces the predicted CO_2 emissions. Each neuron in the output layer performs a weighted sum of its inputs plus a bias term, and the result is passed through the chosen activation function to obtain the final prediction.

3.2 SHAP

The Shapley value originates from cooperative game theory and is used to quantify the contribution of each participant (i.e., "player") to the overall outcome of a collaboration. In this framework, the Shapley value provides a fair allocation mechanism by distributing the total payoff based on the marginal contribution of each player, thereby reflecting each player's importance in the collaboration. This concept has been widely applied in various fields, including the interpretability of machine learning models.

In time series forecasting, due to the non-intuitive nature of time series data, there may be complex dependencies between features, making it difficult to uncover the deep interactions between features by relying solely on the model's output. As a result, there is a growing need for a method that can intuitively understand and quantify the contribution of each feature to the prediction result. The Shapley value, as an optimal contribution measure in game theory, can assign a weight to each feature, reflecting its importance in the prediction outcome. This interpretive approach helps analysts and decision-makers better understand the internal mechanisms and decision-making processes of the model.

In this case, Lundberg and Lee [39] proposed the SHAP method, which leverages Shapley values from game theory to explain complex black-box models. The core idea of SHAP is to explain how a model makes a specific prediction by calculating the marginal contribution of each feature. Specifically, in the SHAP method, each feature value of an sample is interpreted as a "player" in the "game", the deviation between the prediction of a given sample and the average prediction across all samples is considered the "payoff". The essence of the Shapley value is to quantify each feature's marginal contribution and its role in the final prediction. In this way, SHAP assigns a fair contribution value to each feature, revealing the relative importance of each feature in the model's decision-making process.

In machine learning prediction, the "game" refers to the prediction task for a single sample within the dataset. The deviation between the predicted value of an sample and the average predicted value of all samples represents the "payoff" in the game, reflecting the contribution of that sample to the model's prediction. In this game, features act as cooperative participants (i.e., "players"), and their marginal contributions through different subsets of features (coalitions) collectively determine how the total "payoff" is distributed. The Shapley value for each feature reflects its average marginal contribution across all possible feature combinations, ensuring a fair distribution of the total gain.

For non-linear models, feature contributions must be calculated through coalition game theory. The Shapley value is defined by the value function $val_x(S)$ in the context of a feature subset S. The Shapley value for a feature reflects its contribution to the prediction, computed by summing the weighted marginal contributions across all possible feature combinations, as described in Eq. (3), this also reflects the fair distribution principle in coalition games:

$$\phi_i(val) = \sum_{S\subseteq F\setminus\{i\}} \frac{|S|!\,(|F|-|S|-1)!}{|F|!}\left[val(S\cup\{i\}) - val(S)\right] \tag{3}$$

where $F = \{1, \ldots, m\}$ is the full set of features, m represents the number of features and S is any subset of features excluding the i-th feature. The contribution ϕ_i for the i-th feature is obtained by calculating its expected marginal contribution across all possible coalitions. The weight term $\frac{|S|!(|F|-|S|-1)!}{|F|!}$ compensates for the difference in the number of combinations across subsets of different sizes.

The value function of a feature coalition, $val_x(S)$, represents the prediction for the set of features S while marginalizing out the features not in S. It is defined in the form of a conditional expectation integral, as shown in Eq. (4):

$$val_x(S) = \int \hat{C}(x_S, X_{\overline{S}})d\mathbb{P}(X_{\overline{S}}) - E_X(\hat{C}(X)) \tag{4}$$

where x_S denotes the feature values of the sample x in subset S, and $X_{\overline{S}}$ represents the random feature variables in the complement of subset S. The term $\mathbb{P}(X_{\overline{S}})$ is the joint probability distribution over the complementary features.

4 Experiments

4.1 Dataset Preprocessing and Determination of Predicting Objectives

Following the standardized data processing workflow from our previous research, this study integrates panel data from 107 countries spanning from 2000 to 2020 (sample size $N = 2247$), with cross-national joint modeling to enhance the generalizability of the model. The core processing steps include: data cleaning (handling missing values and outliers), feature standardization (using Min-Max Scaling to eliminate dimensionality differences), and driver factor selection (choosing 14 key factors across economic, social, environmental, energy, and technology dimensions). The selected factors are as follows: GDP growth, GDP per capita, Population density, Land area, Latitude, Renewable energy share, Access to electricity, Access to clean fuels for cooking, Electricity from fossil fuels, Electricity from renewables, Low carbon electricity, Primary energy consumption per capita, Energy intensity level of primary energy, Renewable electricity generating capacity per capita. The data split ratio is selected as [Training set:Validation set:Test set] = [6:2:2]. In the implementation, for each valid country–year sample, the 14 driver variables observed over the five-year window are concatenated into a single input vector of length $5 \times 14 = 70$, which serves as the feature vector fed into the MLP.

Similarly, based on the time window selection criteria from our previous research, this study sets a 5-year time window ($W_t = [X_{t-4}, \ldots, X_t] \rightarrow C_{t+1}$) for the following optimization considerations: balancing information richness and complexity (too short a window leads to insufficient feature representation, while too long a window may cause overfitting), dynamic adaptability (capturing the lagging effects of energy policy adjustments and economic growth over the past five years), domain consistency (in many time series forecasting studies, a 5-year window is empirically considered to be an appropriate window size, ensuring data sufficiency while maintaining appropriate flexibility), and prediction performance optimization (the 5-year window strikes a balance between historical data completeness and prediction timeliness).

The 107-country, 20-year panel yields a moderate but diverse sample in which short rolling windows naturally form fixed-length tabular inputs, so that a regularized MLP can capture cross-country nonlinearities while partially smoothing purely idiosyncratic regional effects.

4.2 Quantitative Evaluations

Consistent with our previous research, this study also employs five types of metrics to comprehensively quantify the forecasting performance [40–42]. Let N denote the sample size, C_i and $\hat{C}_i$ represent the actual and predicted values for the i-th sample, respectively, and $\overline{C}$ denotes the mean of the actual values. The detailed description is shown in Table 1.

Table 1: Evaluation metrics and formulas of MLP model

Metric	Formula
R^2 (Coefficient of determination)	$R^2 = 1 - \frac{\sum_{i=1}^{N}(C_i - \hat{C}_i)^2}{\sum_{i=1}^{N}(C_i - \overline{C})^2}$
MSE (Mean Squared Error)	$MSE = \frac{1}{N}\sum_{i=1}^{N}(C_i - \hat{C}_i)^2$
$RMSE$ (Root Mean Squared Error)	$RMSE = \sqrt{MSE}$
MAE (Mean Absolute Error)	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert C_i - \hat{C}_i\rvert$
$MAPE$ (Mean Absolute Percentage Error)	$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left\lvert\frac{C_i - \hat{C}_i}{C_i}\right\rvert \times 100\%$

In addition, the MSE defined in Table 1 is exactly the loss function used during training, i.e., the empirical risk minimization objective in Eq. (1), with $\hat{C}_i = \Phi_{\theta*}(X_i)$ for all evaluation metrics.

4.3 Hyperparameter Selection for Model Optimization

4.3.1 Grid Search for Hyperparameter Tuning

In this study, we employed the Grid Search method [43] to optimize the hyperparameters of the MLP model. Grid search is a technique that exhaustively explores all possible combinations of hyperparameters to determine the optimal hyperparameter configuration. This method systematically traverses a predefined hyperparameter space, evaluates the performance of each combination during model training, and selects the optimal hyperparameters.

The basic principle of grid search is to first define a hyperparameter space, which includes all hyperparameters to be tuned along with their possible values. Then, by traversing all possible parameter combinations in this space, the model is trained and evaluated for each hyperparameter configuration. After each training, we compute the model's performance on the validation set, typically using evaluation metrics such as MSE to measure the model's accuracy and generalization ability. The goal of grid search is to identify the optimal hyperparameter combination that yields the best performance on the validation set, thereby improving the model's predictive performance in real-world applications.

In this study, we constructed a five-dimensional hyperparameter search space that includes key hyperparameters such as hidden layer size, learning rate, batch size, dropout rate, and the number of training epochs. This comprehensive parameter search method helps to minimize the bias associated with manually selecting hyperparameters and enhances the model's generalization capability.

The meanings of the hyperparameters are as follows:

- Hidden layer size: Refers to the number of neurons in each layer. Different combinations of hidden layer sizes allow the model to have varying levels of complexity.

- Learning rate: Controls the magnitude of parameter adjustments during each update. The convergence of the model differs for various learning rates.
- Batch size: Refers to the number of samples selected for each training iteration. Batch size affects the stability of training and computational efficiency, and has a significant impact on the model's final performance.
- Dropout rate: A regularization technique that randomly drops certain nodes in the neural network to prevent overfitting. Different dropout rates control the complexity of the network.
- Number of training epochs: Refers to the number of iterations during model training. The choice of the number of epochs directly impacts the training performance and the risk of overfitting.

Through grid search, we ensured comprehensive optimization of the model across various hyperparameter combinations, eliminating bias from manually selected hyperparameters and improving the model's generalization ability.

4.3.2 Selection of Optimal Hyperparameters

In this study, due to the regression task, the activation function used in each hidden layer of the MLP model is the ReLU function, as shown in Eq. (5). Its purpose is to perform a nonlinear transformation on the input by setting all negative values to zero while retaining positive values. ReLU helps accelerate the training process and reduces the problem of vanishing gradients. The output layer uses a linear activation function by default (i.e., no nonlinear transformation), which is a common practice in regression problems because we want the model to output a continuous value.

$$f(x) = \max(0, x) \tag{5}$$

This study employs the Grid Search method for hyperparameter tuning to determine the optimal configuration of the MLP model. The objective of hyperparameter selection is to enhance the model's training performance and generalization ability, while avoiding overfitting or underfitting. Considering the balance between computational efficiency and model performance, we first constructed a five-dimensional parameter search space as shown in Table 2, covering important parameters such as hidden layer size, learning rate, batch size, dropout rate, and number of epochs.

Table 2: Search space for hyperparameter tuning

Hyperparameter	Search space
Hidden layers	[64, 32, 16], [128, 64, 32], [256, 128, 64], [512, 256, 128]
Learning rate	[0.001, 0.01, 0.1]
Batch size	[16, 32, 64, 128]
Dropout rate	[0.1, 0.2, 0.3, 0.5, 0.6]
Epochs	[20, 50, 70, 100, 150, 200, 300, 500]

Based on the aforementioned hyperparameter search space, a comprehensive hyperparameter tuning process was carried out via grid search, involving 1920 parameter combinations. To prevent overfitting, we used a static validation set (20% of the data) and optimized using *RMSE* as the objective function. The following configuration was ultimately selected as the optimal hyperparameter combination: hidden layer size of [512, 256, 128], learning rate of 0.1, batch size of 128, dropout rate of 0.1, and 50 epochs.

With this optimization approach, after performance evaluation, we achieved the best training results and generalization ability, ensuring the MLP model's stability and accuracy when handling complex data.

In summary, the use of the grid search method not only improved the efficiency and accuracy of hyperparameter tuning but also ensured that the model could achieve optimal predictive performance across different hyperparameter configurations.

4.4 Framework for Scenario Analysis and Model Setup

In this study, we construct multiple policy scenarios based on the impact of different features on CO_2 emissions. We aim to explore how changes in specific factors influence the trend of CO_2 emissions, providing a theoretical foundation for the development of effective emission reduction strategies.

The core of the scenario analysis framework is to simulate the potential impact of changes in various factors on CO_2 emissions by constructing different hypothetical scenarios. Based on the observed influence of features such as Electricity from fossil fuels, Land area and Renewable electricity generating capacity per capita on CO_2 emissions, we develop the scenarios summarised in Table 3. These scenarios include a 50% increase or decrease for each feature, with the changes represented by negative values (−50%) or positive values (+50%) to indicate the variation in the features.

Table 3: Scenario-based feature change overview

Scenario name	Feature	Percentage change
Baseline	No specific feature changes	No change
Scenario 1−	Electricity from fossil fuels	−50%
Scenario 1+	Electricity from fossil fuels	+50%
Scenario 2−	Land area	−50%
Scenario 2+	Land area	+50%
Scenario 3−	Renewable electricity generating capacity per capita	−50%
Scenario 3+	Renewable electricity generating capacity per capita	+50%

The selection of variables used in the scenario analysis is primarily driven by their direct policy relevance and by evidence from the CO_2 emissions and energy-policy literature, rather than by an automatic ranking based on SHAP values. The SHAP analysis provides a complementary perspective when interpreting the scenario results. In Section 5.3, we report global SHAP importance measures and the sign patterns of SHAP values for all 14 drivers. In Section 5.4, these SHAP-based insights are used to explain why, under a given scenario, the predicted emission trajectories move in a particular direction and to check that the simulated policies act through variables that the model itself identifies as influential. Thus, while the scenarios are defined in a policy-driven way, their numerical impacts are interpreted and validated through SHAP-based feature attributions.

Formally, for a given country-year observation with feature vector $\mathbf{x} = (x_1, \ldots, x_{14})$, a scenario acts on a selected feature x_j by applying a relative change

$$x'_j = (1 + \delta)\, x_j, \qquad \delta \in \{-0.5, +0 .5\}, \tag{6}$$

while all other components x_k, $k \neq j$, remain at their historical values. The perturbed input $\mathbf{x}'$ is then passed through the trained MLP to obtain the corresponding counterfactual CO_2 emission. In this study, the value of 50% is used as a symmetric stress test around the historical trajectory to reveal the nonlinear response of the model to substantial, yet interpretable, changes in the selected drivers.

5 Simulation Results and Discussion

5.1 Evaluation of Model Prediction Performance

In this section, we provide a comprehensive evaluation of the model's predictive performance. By systematically analyzing multiple evaluation metrics, including R^2, MSE, $RMSE$, MAE, and $MAPE$, we perform a multidimensional assessment of the model's prediction capabilities.

Before selecting the MLP as the backbone model, we also conducted a preliminary comparison with several widely used tree-based machine learning methods (Random Forest, XGBoost, LightGBM and CatBoost) on the same multi-country panel dataset and using the same evaluation metrics. The aggregated results are summarized in Table 4. The Avg_R^2 values of these baseline models range between 0.8919 and 0.9290, with Avg_RMSE on the order of 2.7×10^5–3.4×10^5 and Avg_MAE around 5.4×10^4–7.4×10^4, which are all clearly inferior to the performance of the proposed MLP model reported in Table 5. Within the set of models tested in this study, these preliminary results indicate that the MLP provides the best trade-off between predictive accuracy and numerical stability, and therefore it is chosen as the basis for the subsequent SHAP analysis and policy scenario simulations.

Table 4: Aggregated performance of tree-based baseline models on the multi-country CO_2 emissions dataset

Model	Avg_R^2	Avg_MSE	Avg_RMSE	Avg_MAE	Avg_MAPE
Random Forest	0.8919	113,337,578,108.7750	336,656.4690	73,959.4302	2022.2320
XGBoost	0.8936	111,403,327,458.1250	333,771.3700	53,801.3430	132.2520
LightGBM	0.9232	80,463,645,220.4352	283,661.1451	57,984.4549	271.1507
CatBoost	0.9290	74,316,642,461.0289	272,610.7893	60,572.9513	2108.9588

Table 5: Quantitative performance metrics of MLP

Metric	MLP
Avg_R^2	0.9951
Avg_MSE	5,131,813,424.0152
Avg_$RMSE$	71,636.6765
Avg_MAE	27,063.5670
Avg_$MAPE$	764.0848

Experimental data indicate that the MLP model demonstrates significant advantages in nonlinear fitting, showcasing its strong predictive power. The average R^2 value reaches 0.9951, meaning the model can explain approximately 99.51% of the data variability, which reflects its high fitting accuracy in capturing complex relationships and patterns in the data.

In terms of error control, the MLP model also performs exceptionally well. The average MSE is 5,131,813,424.0152, and the average $RMSE$ is 71,636.6765, indicating that the model has outstanding performance in reducing prediction errors and can fit the actual data with a high level of accuracy. Additionally, the average MAE remains stable at 27,063.5670, demonstrating the model's strong robustness in terms of absolute error, maintaining stable performance under varying prediction conditions.

However, it is worth noting that despite the MLP model's excellent performance across most metrics, the average $MAPE$ is as high as 764.0848%. This result suggests that the model may exhibit significant outlier behavior in some cases. The relatively high $MAPE$ value indicates that certain data points might cause large deviations due to extreme conditions or noisy data. Therefore, further research and treatment of these outliers are needed to improve the model's stability and accuracy across all scenarios in practical applications. Given that CO_2 emissions in some small economies and in the early years of the study period are very low compared with major emitters, even moderate absolute errors can lead to extremely large relative errors. Therefore, the very large $MAPE$ value mainly reflects the numerical sensitivity of this metric in low-emission regimes rather than a general lack of fit of the model.

In conclusion, the MLP model performs exceptionally well across most evaluation metrics, especially in terms of nonlinear fitting capability and error control. However, the presence of outliers points to areas for further optimization and adjustment, particularly in handling extreme data conditions.

5.2 Learning Curve for MLP Model

In this section, we present the learning curves of the MLP model, showing how the R^2 and MSE evolve with the number of training iterations (epochs).

Firstly, from the R^2 learning curve (Fig. 2), we observe that as the number of training epochs increases, the R^2 on the training set continuously rises and eventually stabilizes, approaching 1.0. This indicates that the model's predictive capability on the training dataset significantly improves. The R^2 on the validation dataset also shows a gradual upward trend, although there are some fluctuations during the process, it ultimately stabilizes. Notably, the model achieved an average R^2 of 0.9951 on the test set, demonstrating that the MLP model exhibits excellent stability and generalization ability during training, effectively predicting unseen data and showcasing its reliability for practical applications.

Next, from the MSE learning curve (Fig. 3), we see that as the number of training epochs increases, the training error significantly decreases and eventually levels off. While there is some fluctuation in the validation error, the overall trend stabilizes. The average MSE on the test set is 5,131,813,424.0152, reflecting that the MLP model effectively controlled the error during training, validating its strong performance in error control and optimization.

Overall, despite some fluctuations during training, the MLP model ultimately demonstrates strong performance in both R^2 and MSE, validating its reliability and accuracy for complex prediction tasks.

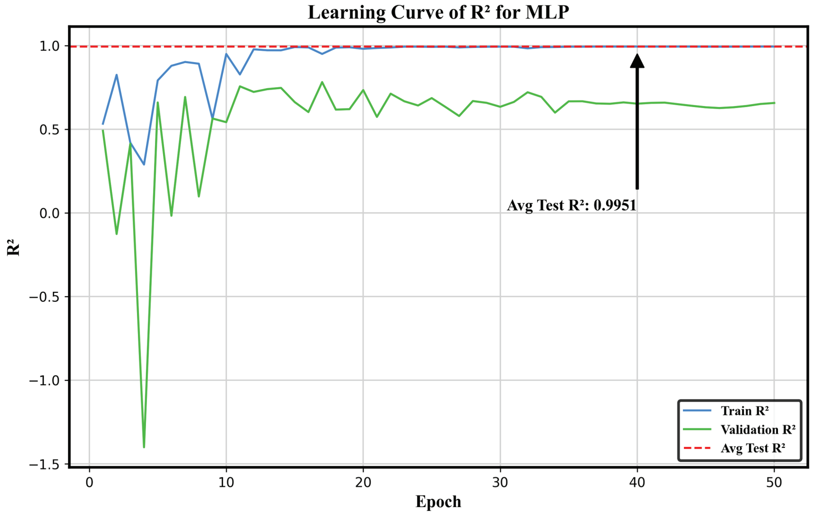

Figure 2: Learning curve of R^2 for MLP

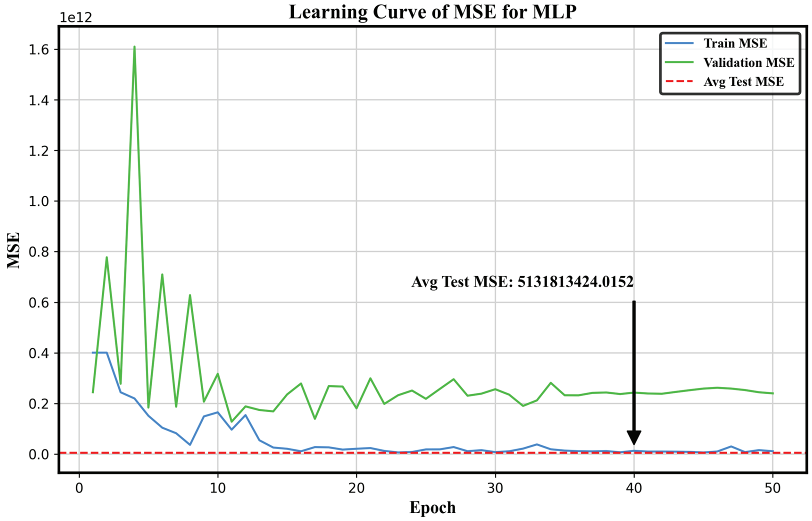

Figure 3: Learning curve of MSE for MLP

5.3 Evaluation of Feature Contributions Using SHAP

In this section, we will use the SHAP method to evaluate the feature contributions in the model. The core theory of SHAP is derived from Shapley values in game theory, which aims to fairly allocate the marginal contributions of multiple features in the model's output, thus achieving interpretability in model predictions. By analyzing various SHAP plots, we can gain a deeper understanding of how the model makes predictions regarding CO_2 emissions based on different features. It is worth emphasising that Figs. 4 and 5 report global behaviour across all country-year samples, while local, instance-specific feature attributions are analysed separately in the SHAP waterfall plot in Fig. 6.

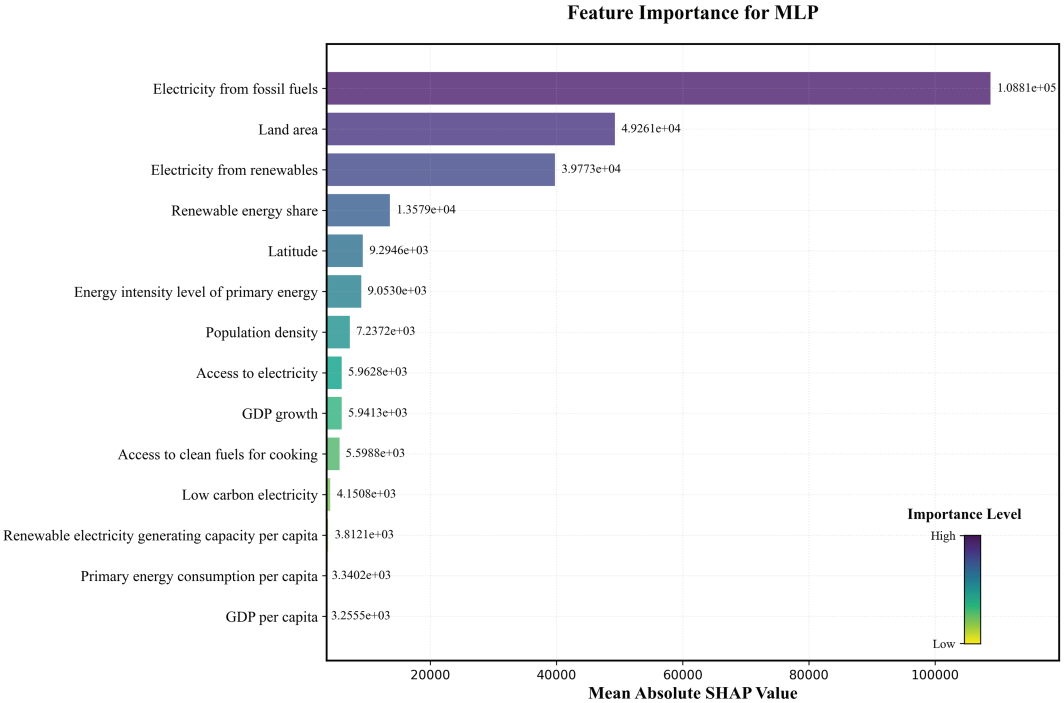

Figure 4: Feature importance for MLP

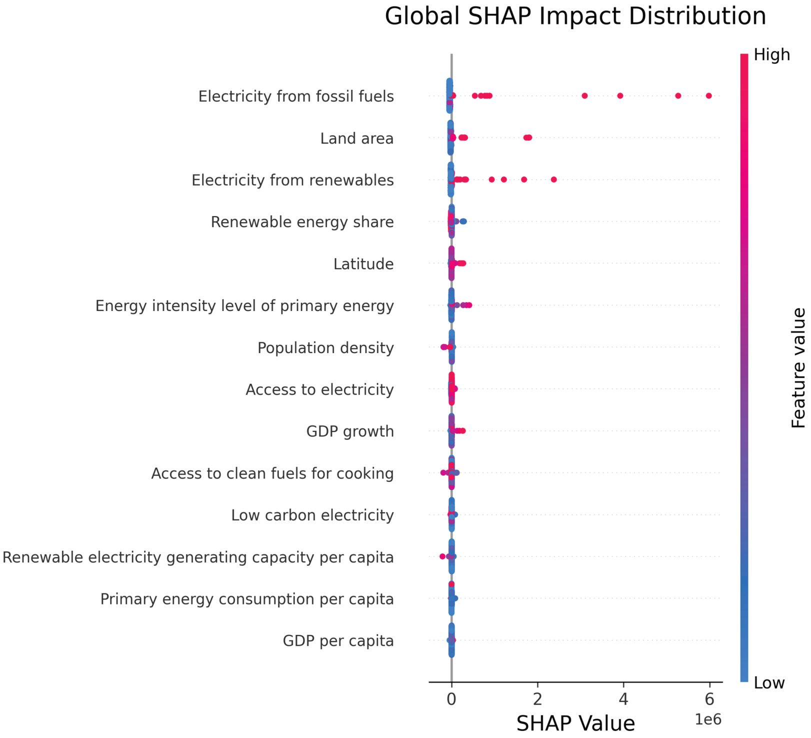

Figure 5: Global SHAP impact distribution for MLP

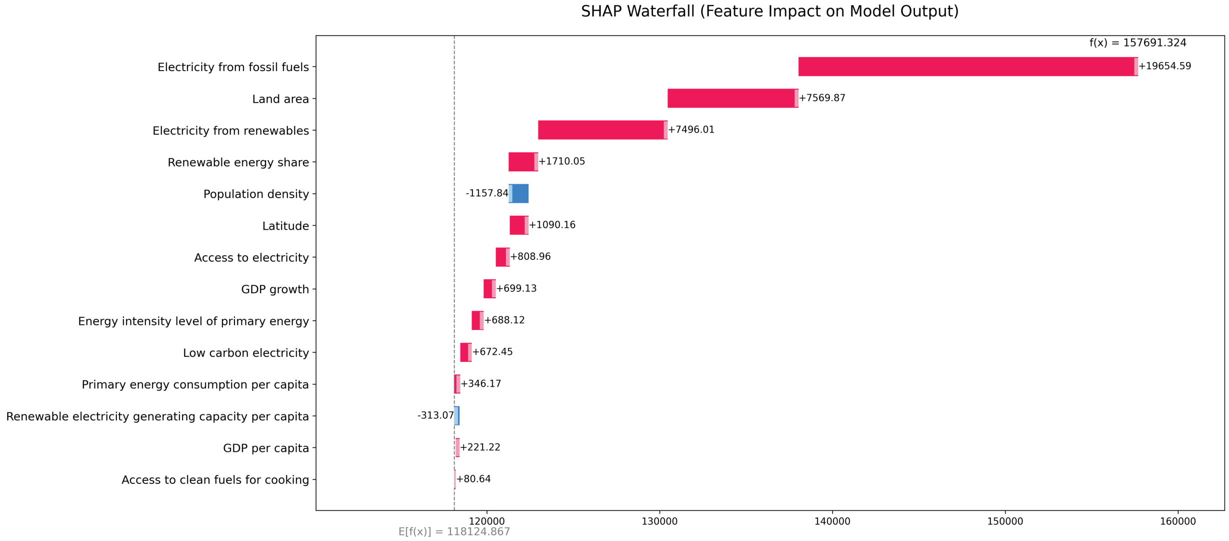

Figure 6: SHAP Waterfall for MLP

5.3.1 Feature Importance for MLP

The feature importance plot of the MLP model (Fig. 4) uses a SHAP-based feature importance ranking method, showing the global importance ranking of each feature in carbon emission prediction. Through the quantitative assessment of feature importance, it reveals the priority differences of key driving factors.

The plot employs a multi-dimensional visualization method, clearly demonstrating the importance of each feature. The horizontal axis (X-axis) displays the contribution value of each feature, and the length of the rectangular bars intuitively reflects the relative importance of each feature. The vertical axis (Y-axis) lists all the feature variables involved in the analysis. A color gradient from dark to light is used for color coding, where darker colors represent greater contributions to the model's prediction, and lighter colors indicate smaller contributions. The gradient effect of colors effectively reflects the hierarchical distribution of feature importance. Furthermore, the importance values are represented in scientific notation, the larger the value, the greater the influence of that feature on the model's prediction.

From Fig. 4, it can be seen that "Electricity from fossil fuels" has the largest impact on the model, with a SHAP value of 1.0881e+05, significantly higher than other features. This indicates that the use of fossil fuels plays a crucial role in the prediction results. This feature's importance may reflect the widespread use of fossil fuels and the central role of energy structure in the model. "Land area" follows closely with a SHAP value of 4.9261e+04, also showing a strong influence. As a natural geographic feature, land area may affect energy demand, infrastructure development, and other factors, thus holding significant weight in the model. "Electricity from renewables" and "Renewable energy share" rank third and fourth, with SHAP values of 3.9773e+04 and 1.3579e+04, respectively. This indicates that while the importance of renewable energy in the global energy transition is increasing, it still holds a secondary position compared to fossil fuel electricity.

The SHAP value for "Latitude" is 9.2946e+03, indicating that geographic location may influence energy demand and supply in certain regions. "Energy intensity level of primary energy" follows with a SHAP value of 9.0530e+03, which may be related to energy efficiency and environmental impact. “Population density” and “Access to electricity” rank sixth and seventh, with SHAP values of 7.2372e+03 and 5.9628e+03, respectively. Population density may affect energy demand, while access to electricity reflects the extent of electricity supply coverage in different regions. In developing areas, regions with lower access to electricity may rely more on traditional or alternative energy sources, which could impact energy structure and consumption patterns. “GDP growth” has a SHAP value of 5.9413e+03, indicating that this socioeconomic factor somewhat influences energy consumption and CO_2 emission patterns. “Access to clean fuels for cooking” has a SHAP value of 5.5988e+03, indicating that the share of renewable energy also affects CO_2 emissions, reflecting the significant role of clean energy in the energy structure.

Other features in Fig. 4, such as “Low carbon electricity” (4.1508e+03), “Renewable electricity generating capacity per capita” (3.8121e+03), “Primary energy consumption per capita,” and “GDP per capita,” while ranked lower, still have an impact on the model prediction. Particularly, “Low carbon electricity” shows a more noticeable role in scenarios considering low-carbon energy policies or green energy. “Renewable electricity generating capacity per capita” demonstrates the importance of renewable electricity generation capacity per person on the model, especially in assessing energy distribution and sustainability.

Overall, the model’s prediction results are mainly influenced by energy-related features, especially those related to fossil fuels and renewable energy, which aligns with the current global energy transition and climate change issues. Land area and geographic features also play an important role in the model, which may be related to regional differences in energy demand and supply.

5.3.2 Global SHAP Impact Distribution for MLP

The SHAP feature importance scatter plot generated based on the MLP model (Fig. 5) presents global feature importance, revealing the heterogeneous impact of each feature on the CO_2 emissions prediction model through multidimensional data mapping. The x-axis of the plot shows the SHAP values, which represent the marginal contribution direction and intensity of each feature to the model’s output (positive values drive an increase in CO_2 emissions predictions, while negative values suppress the prediction results). The y-axis ranks the features in descending order of their average absolute SHAP values, and the color mapping reflects the normalized degree of the feature’s original values (red: high values, blue: low values). The width of the scatter points indicates the sample distribution density.

From the results shown in Fig. 5, it is evident that different features have significantly varying marginal contributions (driving effects on CO_2 emissions) to the model output, and these contributions exhibit complex distribution patterns in different groups. Among these features, “Electricity from fossil fuels,” “Land area,” and “Electricity from renewables” have the highest average absolute SHAP values, suggesting that they have the most prominent global explanatory power on CO_2 emissions predictions. Further analysis reveals that the SHAP values for these features are predominantly in the positive range of the x-axis and concentrated in the higher SHAP value range, indicating that higher feature values significantly increase the model’s CO_2 emissions predictions. These observations highlight that the model heavily relies on energy consumption (especially the proportion of fossil fuel and renewable energy) and land area when predicting CO_2 emissions, underscoring their importance in the prediction process.

However, the analysis of feature value group heterogeneity further reveals that the impact patterns of different features on CO_2 emissions vary. The SHAP values for “Access to electricity,” “Access to clean fuels for cooking," "Low carbon electricity," "Primary energy consumption per capita," and "GDP per capita" show a bidirectional distribution, indicating that these features have varying impacts on CO_2 emissions across different feature value ranges. In some cases, they may drive up emissions, while in others, they might have a reduction effect.

Additionally, although features such as "Renewable energy share," "Latitude," "Energy intensity level of primary energy," and "GDP growth" mostly exhibit bidirectional distributions of SHAP values, a small proportion of high-normalized values (red scatter points) are concentrated in the positive SHAP range, suggesting that higher values of these features significantly drive CO_2 emissions upward. Conversely, "Population density" and "Renewable electricity generating capacity per capita" exhibit the opposite trend, with high-normalized values (red scatter points) predominantly concentrated in the negative SHAP range, indicating that higher values of these features are associated with lower CO_2 emissions.

The sample density indicated by the scatter point width shows that the impact of most features is concentrated in the lower SHAP value range (narrow width areas), while the extreme value regions (such as the high-density positive distribution of "Electricity from fossil fuels") highlight the model's sensitivity to a few core variables.

5.3.3 SHAP Waterfall for MLP

The SHAP waterfall plot based on the MLP model (Fig. 6) illustrates how each feature contributes to the predicted CO_2 emissions for one representative country–year observation. Features are ordered by the magnitude of their local contributions for this sample; red bars increase the prediction relative to the base value, while blue bars decrease it.

The $f(x)$ value of 157,691.32 in Fig. 6 represents the model's final predicted output for this particular sample. Among the features, "Electricity from fossil fuels," "Land area," and "Electricity from renewables" make the largest positive contributions to the global prediction, with contribution values of +19,654.59, +7569.87, and +7496.01, respectively. These larger contributions indicate that electricity sources and land resources are key driving factors in the model's prediction and are associated with an increase in CO_2 emissions. Features such as "Population density," "Renewable electricity generating capacity per capita," and "Latitude" have a negative impact, suggesting that higher values of these features are linked to a reduction in CO_2 emissions.

Additionally, several other features have relatively smaller contributions. For the local waterfall plot in Fig. 6, the contribution of "Renewable energy share" is positive (for example, +1710.05 in the reported scale). This positive SHAP value does not imply that a higher share of renewables is intrinsically emission-increasing; rather, it reflects the association learned by the MLP model under the joint influence of all covariates. In our dataset, countries with a large renewable share are often high-demand systems in a transition phase, where renewables currently complement rather than fully replace fossil generation. The SHAP result should therefore be interpreted as a model-based association conditioned on the remaining features, not as direct causal evidence about renewable policies. "Latitude" contributes +1090.16, showing a positive impact of geographic location on CO_2 emissions. "Access to electricity" has a contribution of +808.96, where areas with better electricity access tend to have higher energy consumption and CO_2 emissions, especially in fossil fuel-dominated power systems. However, areas with insufficient electricity access might have lower energy consumption and, consequently, lower CO_2 emissions. "GDP growth" contributes +699.13, as economic growth generally drives increased energy consumption, particularly fossil fuels, leading to higher CO_2 emissions. The contribution of "Energy intensity level of primary energy" is +688.12, where higher energy intensity usually implies lower energy efficiency, meaning more energy consumption is required for the same economic activity, leading to more CO_2 emissions. “Low carbon electricity” contributes +672.45, where, in theory, low-carbon electricity reduces CO_2 emissions. However, its application may be limited due to infrastructure constraints or insufficient replacement of fossil fuel power, which might still lead to higher emissions. “Primary energy consumption per capita” contributes +346.17, indicating that higher energy consumption per person generally leads to increased CO_2 emissions, especially in fossil fuel-reliant energy systems. “GDP per capita” contributes +221.22, where higher per capita GDP generally correlates with higher living standards and energy consumption, particularly in developed or high-income regions, driving higher CO_2 emissions. “Access to clean fuels for cooking” contributes +80.64, where the availability of clean fuels helps reduce CO_2 emissions. However, in some regions, limited access to clean fuels might still lead to the use of traditional energy sources, reducing the potential impact on CO_2 emissions.

In summary, the SHAP-based analysis highlights the importance of each feature in the CO_2 emissions prediction model, particularly through the integration of key indicators from economic, social, environmental, energy, and technology domains. Among these, energy and social factors play a crucial role in determining CO_2 emissions.

5.4 Evaluation of Prediction Performance in Various Scenarios

This study integrates a scenario analysis framework by constructing six distinct scenarios to simulate the impact of various features on global CO_2 emission trends. The results are presented in Figs. 7–9. According to the quantitative evaluation in the previous subsection, the proposed MLP model attains a coefficient of determination of $R^2 = 0.9951$ and a low RMSE on the test set. These results indicate that the network is able to approximate the highly non-linear relationships between the 14 driving factors and national CO_2 emissions with high fidelity. Therefore, the multi-policy scenarios discussed below can be regarded as numerically meaningful and robust, since the simulated trajectories are generated by a model that accurately captures the main interaction patterns in the data.

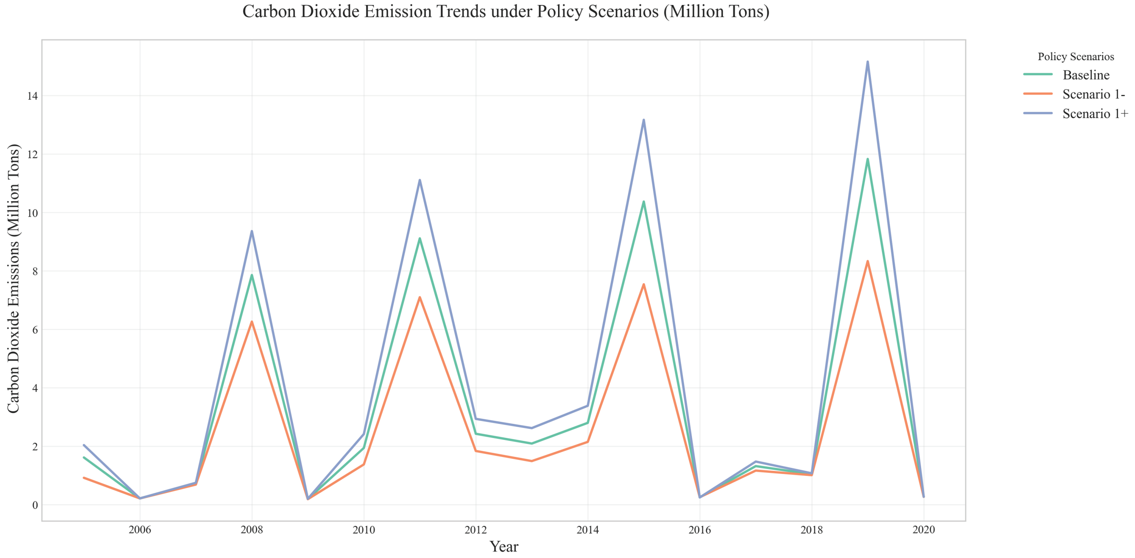

Figure 7: Trends under policy scenario for electricity from fossil fuels

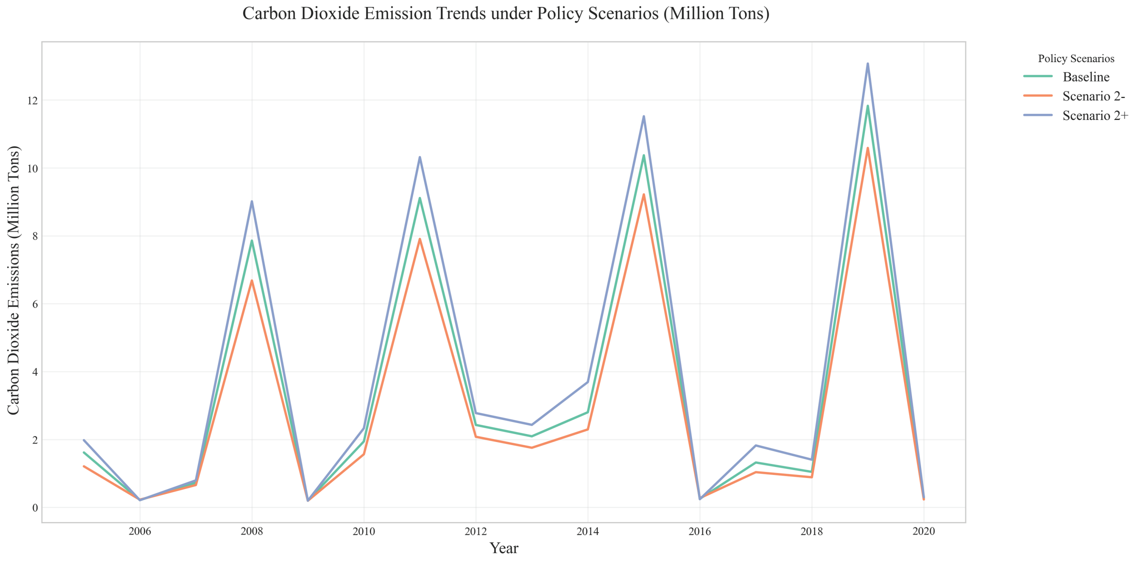

Figure 8: Trends under policy scenario for land area

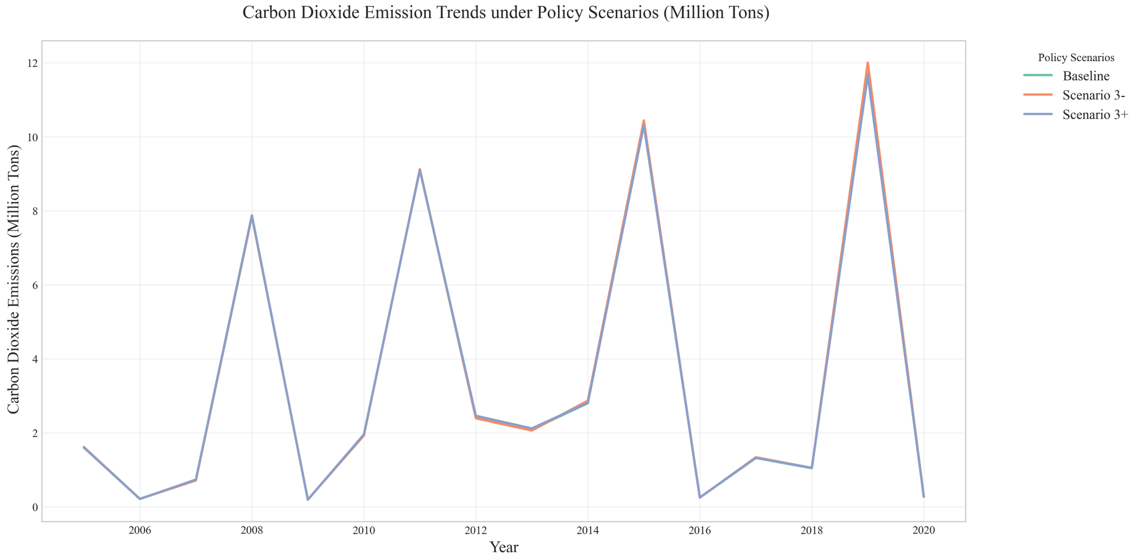

Figure 9: Trends under policy scenario for renewable electricity generating capacity per Capita

As shown in Fig. 7, when Electricity from fossil fuels is reduced by 50% (Scenario 1−), CO_2 emissions exhibit a significant downward trend between 2006 and 2020. In contrast, a 50% increase in Electricity from fossil fuels (Scenario 1+) leads to a sharp rise in CO_2 emissions. This indicates that the reliance on fossil fuels is a core driver of CO_2 emissions, and reducing their share has a direct effect on emissions reduction.

As shown in Fig. 8, when the Land area is reduced by 50% (Scenario 2−), CO_2 emissions slightly decrease in the short term. Conversely, when the Land area is increased by 50% (Scenario 2+), it is likely associated with higher land use efficiency, which leads to a small increase in CO_2 emissions.

This result highlights the dual role of land management in carbon neutrality goals: poor land development may exacerbate emissions, while ecological protection can enhance carbon sequestration.

As shown in Fig. 9, reducing Renewable electricity generating capacity per capita (Scenario 3−) leads to an increase in CO_2 emissions. This is because a reduction in this factor results in insufficient renewable energy supply, and electricity demand cannot be fully met by clean energy, forcing reliance on traditional fossil fuels (such as coal and gas) to fill the gap. Since these traditional energy sources have high CO_2 emissions, this leads to a significant rise in CO_2 emissions. When Renewable electricity generating capacity per capita is increased (Scenario 3+), CO_2 emissions show a downward trend. This is because, with an increase in this factor, countries or regions can meet electricity demand with more low-carbon energy sources such as wind and solar, reducing dependence on fossil fuels. This shift directly leads to a reduction in CO_2 emissions in the power generation sector, as renewable energy sources emit less CO_2 during power generation.

This study reveals the dynamic impact of different policy interventions on global CO_2 emissions through multi-scenario simulations. At the same time, the present scenario design has clear limitations: each factor is perturbed in isolation while all other drivers are kept at their historical values, and the $\pm 50\%$ changes are stylised stress tests rather than policy targets calibrated to specific national or sectoral plans. The trajectories in Figs. 7–9 should therefore be interpreted as local sensitivity experiments around the historical path, illustrating the qualitative response of the model to substantial variations in key drivers rather than fully specified long-term mitigation pathways. In summary, achieving carbon neutrality requires a multidimensional and coordinated strategy: in the short term, priority should be given to reducing the share of fossil fuels, in the medium term, enhancing the penetration of renewable energy technologies and improving grid infrastructure, and in the long term, optimizing emissions reduction pathways through land ecological restoration policies. Policy formulation should avoid relying on a single factor and instead coordinate technological substitutions, infrastructure investments, and institutional innovations to systematically address the complex dynamics of CO_2 emissions. Unlike earlier CO_2 forecasting studies that employ deep neural networks or traditional machine-learning regressors as opaque "black-box" models, our framework explicitly couples a high-capacity MLP with SHAP-based explanations. This "DL + XAI" paradigm preserves the non-linear predictive power of deep learning while providing transparent, feature-level attribution that directly supports the design and interpretation of the emission-reduction scenarios.

6 Conclusion

This study explores the effectiveness of using DL frameworks, particularly the representative MLP model, in predicting CO_2 emissions. It integrates XAI techniques, specifically using SHAP analysis, to successfully quantify the contribution of each feature to the prediction results. The study not only provides predictions of CO_2 emissions for different countries globally across various years using DL models but also offers more transparent tools that make the decision-making process of the "black-box" model understandable, further enhancing the model's credibility in practical applications.

The MLP model excels at capturing the high-dimensional nonlinear relationships between driving factors, making predictions more accurate when simulating complex policy scenarios. The model exhibits excellent global predictive performance in terms of R^2 and RMSE (and MAE), while the very large MAPE value highlights a limitation of relative-error metrics for low-emission observations and should be borne in mind when interpreting country-level results.

SHAP provides a way to clearly display the contribution of each feature to the model's output, helping decision-makers understand the role of different features in the prediction. SHAP-based explainability analysis reveals that energy-related features, particularly “Electricity from fossil fuels” and “Electricity from renewables,” play a dominant role in the model’s predictions. Specifically, fossil fuel use contributes the most to emissions, while a higher proportion of renewable energy helps reduce emissions. Additionally, social factors such as “Land area” have a significant impact on the model, potentially related to energy demand, infrastructure development, and regional energy supply-demand differences. Geographic factors, like “Latitude,” also affect CO_2 emissions due to differing energy consumption patterns and economic development levels across regions. These findings align with the global challenges of energy transition and emphasize the key role of optimizing energy structures in emission reduction.

This “DL + XAI” integrated analysis approach successfully addresses the challenge of model explainability. Unlike purely black-box deep-learning models and less expressive conventional ML baselines, it combines high-fidelity non-linear forecasting with explicit feature-level attributions that can be directly used to support numerical scenario design and policy analysis. It not only deepens our theoretical understanding of the multi-scale driving mechanisms of CO_2 emissions but also helps policymakers better comprehend the complex relationships between multi-dimensional driving factors and CO_2 emissions. This approach provides quantitative support for formulating differentiated climate governance strategies and can be applied to decision support across various policy scenarios.

Through the multi-policy scenario analysis framework proposed in this study, we demonstrate the potential impact of changes in different features (such as “Electricity from fossil fuels,” “Land area,” and “Renewable electricity generating capacity per capita”) on CO_2 emission trends. Scenario simulations further validate the dynamic impact of policy interventions. The simulation results suggest that reducing fossil fuel use in the short term can significantly lower emissions, while increasing the proportion of renewable energy contributes to achieving long-term emission reduction targets. Additionally, land ecological restoration policies play an important role in the long term. Therefore, achieving carbon neutrality requires a comprehensive consideration of multi-dimensional policy measures, avoiding reliance on a single policy, and coordinating aspects such as energy substitution, infrastructure construction, and institutional innovation to systematically address the complex dynamics of CO_2 emissions.

The contribution of this study lies not only in combining the existing MLP model with SHAP for CO_2 emission predictions, successfully achieving a collaborative optimization of prediction accuracy and model explainability, but also in providing policymakers with different policy scenario analyses, promoting the design of more scientific and sustainable climate governance strategies. Future research could further optimize the model and explore additional policy scenarios, as well as investigate the long-term impacts of different feature combinations on CO_2 emissions, particularly in the context of global climate change. This could include optimizing energy policies, enhancing the use of renewable energy, and reducing reliance on fossil fuels to advance the global carbon neutrality process.

In addition, our macro-level DL + XAI framework can be extended to sector-specific, high-consumption infrastructures such as data centre networks. Recent work by Du et al. [44] shows that decarbonisation of data centre networks can be achieved through computing power migration and coordinated use of energy storage. Integrating such operational strategies with our emission forecasts would enable multi-scale planning from national climate targets down to ICT-intensive systems. Moreover, future extensions could embed the scenario module into a stochastic optimisation framework that explicitly separates exogenous and endogenous (decision-dependent) uncertainty, following the ideas of Giannelos et al. [45], thereby yielding more robust long-term emission planning under uncertain technology adoption and demand-side response.

Acknowledgement: The authors did not receive support from any organization for the submitted work.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: All authors contributed to the study conception and design. Material preparation, data collection, experiments, and analysis were performed by Ruimin Ma, Qiong Li and Alexander Kovshov. The first draft of the manuscript was written by Ruimin Ma. And all authors commented on previous versions of the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Tanwar, A. Global data on sustainable energy (2000–2020); 2023. Kaggle. Available from: https://www.kaggle.com/dsv/6327347 (accessed on 01 December 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Narayan PK, Narayan S. Carbon dioxide emissions and economic growth: panel data evidence from developing countries. Energy Policy. 2010;38(1):661–6. doi:10.1016/j.enpol.2009.09.005.
2. Gür TM. Carbon dioxide emissions, capture, storage and utilization: review of materials, processes and technologies. Prog Energy Combust Sci. 2022;89:100965. doi:10.1016/j.pecs.2021.100965.
3. Mardani A, Liao H, Nilashi M, Alrasheedi M, Cavallaro F. A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. J Clean Prod. 2020;275:122942. doi:10.1016/j.jclepro.2020.122942.
4. Kadam P, Vijayumar S. Prediction model: CO_2 emission using machine learning. In: 2018 3rd International Conference for Convergence in Technology (I2CT); 2018 Apr 6–8; Pune, India. p. 1–3.
5. Qin J, Gong N. The estimation of the carbon dioxide emission and driving factors in China based on machine learning methods. Sustainable Production Consump. 2022;33:218–29. doi:10.1016/j.spc.2022.06.027.
6. As M, Bilir T. Machine learning algorithms for energy efficiency: mitigating carbon dioxide emissions and optimizing costs in a hospital infrastructure. Energy Build. 2024;318:114494. doi:10.1016/j.enbuild.2024.114494.
7. AlKheder S, Almusalam A. Forecasting of carbon dioxide emissions from power plants in Kuwait using United States Environmental Protection Agency, Intergovernmental panel on climate change, and machine learning methods. Renewable Energy. 2022;191:819–27. doi:10.1016/j.renene.2022.04.023.
8. Li S, Siu YW, Zhao G. Driving factors of CO_2 emissions: further study based on machine learning. Front Environ Sci. 2021;9:721517. doi:10.3389/fenvs.2021.721517.
9. Wang Y, Wang R, Tanaka K, Ciais P, Penuelas J, Balkanski Y, et al. Global spatiotemporal optimization of photovoltaic and wind power to achieve the Paris Agreement targets. Nat Commun. 2025;16(1):1–19. doi:10.1038/s41467-025-57292-w.
10. Luo Z, He T, Lv Z, Zhao J, Zhang Z, Wang Y, et al. Insights into transportation CO_2 emissions with big data and artificial intelligence. Patterns. 2025;6(4):101186. doi:10.1016/j.patter.2025.101186.
11. Sharma SS. Determinants of carbon dioxide emissions: empirical evidence from 69 countries. Appl Energy. 2011;88(1):376–82. doi:10.1016/j.apenergy.2010.07.022.

12. Fan A, Xiong Y, Yan J, Yang L, Shu Y, Chen J. Microscopic characteristics and influencing factors of ship emissions based on onboard measurements. Transp Res Part D Transp Environ. 2024;133:104300. doi:10.1016/j.trd.2024.104300.
13. Khan MK, Teng JZ, Khan MI, Khan MO. Impact of globalization, economic factors and energy consumption on CO_2 emissions in Pakistan. Sci Total Environ. 2019;688:424–36. doi:10.1016/j.scitotenv.2019.06.065.
14. Zhang Q, Boersma KF, Zhao B, Eskes H, Chen C, Zheng H, et al. Quantifying daily NO_x and CO_2 emissions from Wuhan using satellite observations from TROPOMI and OCO-2. Atmospheric Chem Phys Discuss. 2022;2022:1–18. doi:10.5194/acp-23-551-2023.
15. Wu R, Xie Z, Wang J, Wang S. Estimating the environmental Kuznets curve and its influencing factors of CO_2 emissions: insights from development stages and rebound effects. Appl Geogr. 2025;174:103475. doi:10.1016/j.apgeog.2024.103475.
16. Zennaro F, Furlan E, Simeoni C, Torresan S, Aslan S, Critto A, et al. Exploring machine learning potential for climate change risk assessment. Earth-Sci Rev. 2021;220:103752. doi:10.1016/j.earscirev.2021.103752.
17. Li H, Yang Y, Wang H, Wang P, Yue X, Liao H. Projected aerosol changes driven by emissions and climate change using a machine learning method. Environ Sci Technol. 2022;56(7):3884–93.
18. Zhong S, Zhang K, Bagheri M, Burken JG, Gu A, Li B, et al. Machine learning: new ideas and tools in environmental science and engineering. Environ Sci Technol. 2021;55(19):12741–54.
19. Hamdan A, Ibekwe KI, Etukudoh EA, Umoh AA, Ilojianya VI. AI and machine learning in climate change research: a review of predictive models and environmental impact. World J Adv Res Rev. 2024;21(1): 1999–2008. doi:10.30574/wjarr.2024.21.1.0257.
20. Martiny A. Towards sustainable AI: Monitoring and analysis of carbon emissions in machine learning algorithms. Torino, Italy: Politecnico di Torino; 2023.
21. Rolnick D, Donti PL, Kaack LH, Kochanski K, Lacoste A, Sankaran K, et al. Tackling climate change with machine learning. ACM Comput Surv (CSUR). 2022;55(2):1–96. doi:10.1145/3485128.
22. Wang W, Li D, Zhou S, Wang Y, Yu L. Exploring the key influencing factors of low-carbon innovation from urban characteristics in China using interpretable machine learning. Environ Impact Assess Rev. 2024;107:107573. doi:10.1016/j.eiar.2024.107573.
23. AlShafeey M, Rashdan O. Quantifying the impact of energy consumption sources on GHG emissions in major economies: a machine learning approach. Energy Strategy Rev. 2023;49:101159. doi:10.1016/j.esr.2023.101159.
24. Zhang Y, Zhao S, Kazarinov N, Petrov YV. Neural networks with iterative parameter generation for determining parameters of constitutive models. Cybern Phys. 2024;13:334–45.
25. Ladi T, Jabalameli S, Sharifi A. Applications of machine learning and deep learning methods for climate change mitigation and adaptation. Environ Plan B Urban Anal City Sci. 2022;49(4):1314–30. doi:10.1177/23998083221085281.
26. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44. doi:10.1038/nature14539.
27. Zhang Y, Logachov A, Smirnov A, Kazarinov N. Artificial neural network surrogates for impact problem simulations. Cybern Phys. 2025;14(2):200–13.
28. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: a review. Neurocomputing. 2016;187:27–48. doi:10.1016/j.neucom.2015.09.116.
29. Zhang Y, Logachev A, Smirnov A, Kazarinov N. Artificial neural networks for impact strength prediction of composite barriers. Materials. 2025;18(13):3001. doi:10.3390/ma18133001.
30. Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N. Multilayer perceptron and neural networks. WSEAS Trans Circ Syst. 2009;8(7):579–88.
31. Singh J, Banerjee R. A study on single and multi-layer perceptron neural network. In: Proceedings of the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC); 2019 Mar 27–29; Erode, India. p. 35–40.

32. Nohara Y, Matsumoto K, Soejima H, Nakashima N. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput Methods Programs Biomed. 2022;214:106584. doi:10.1016/j.cmpb.2021.106584.
33. Walia S, Kumar K, Agarwal S, Kim H. Using XAI for deep learning-based image manipulation detection with shapley additive explanation. Symmetry. 2022;14(8):1611. doi:10.3390/sym14081611.
34. Ullah I, Liu K, Yamamoto T, Zahid M, Jamal A. Prediction of electric vehicle charging duration time using ensemble machine learning algorithm and Shapley additive explanations. Int J Energy Res. 2022;46(11):15211–30. doi:10.1002/er.8219.
35. García MV, Aznarte JL. Shapley additive explanations for NO_2 forecasting. Ecol Inform. 2020;56:101039. doi:10.1016/j.ecoinf.2019.101039.
36. Zhang J, Ma X, Zhang J, Sun D, Zhou X, Mi C, et al. Insights into geospatial heterogeneity of landslide susceptibility based on the SHAP-XGBoost model. J Environ Manage. 2023;332:117357. doi:10.1016/j.jenvman.2023.117357.
37. Tanwar A. Global Data on Sustainable Energy (2000–2020). Kaggle. [cited 2025 Dec 1]. Available from: https://www.kaggle.com/dsv/6327347.
38. Wu L, Liu S, Liu D, Fang Z, Xu H. Modelling and forecasting CO_2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multi-variable grey model. Energy. 2015;79:489–95. doi:10.1016/j.energy.2014.11.052.
39. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017 Dec 4–9; Long Beach, CA, USA. p. 4768–77.
40. Li X, Zhang X. A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China. Environ Sci Pollution Res. 2023;30(55):117485–502. doi:10.1007/s11356-023-30428-5.
41. Faruque MO, Rabby MAJ, Hossain MA, Islam MR, Rashid MMU, Muyeen S. A comparative analysis to forecast carbon dioxide emissions. Energy Rep. 2022;8:8046–60. doi:10.1016/j.egyr.2022.06.025.
42. Namboori S. Forecasting carbon dioxide emissions in the United States using machine learning. Dublin, Ireland: National College of Ireland; 2020.
43. Syarif I, Prugel-Bennett A, Wills G. SVM parameter optimization using grid search and genetic algorithm to improve classification performance. TELKOMNIKA (Telecommun Comput Electron Control). 2016;14(4):1502–9. doi:10.12928/telkomnika.v14i4.3956.
44. Du Z, Yin H, Zhang X, Hu H, Liu T, Hou M, et al. Decarbonisation of data centre networks through computing power migration. In: Proceedings of the 2025 IEEE 5th International Conference on Computer Communication and Artificial Intelligence (CCAI); 2025 May 23–25; Haikou, China. p. 871–6.
45. Giannelos S, Konstantelos I, Zhang X, Strbac G. A stochastic optimization model for network expansion planning under exogenous and endogenous uncertainty. Elect Power Syst Res. 2025;248:111894. doi:10.1016/j.epsr.2025.111894.