Machine Learning-Based Modeling of the Environmental Degradation, Institutional Quality, and Economic Growth

Sami Ben Jabeur[1] · Houssein Ballouk[2] · Wissal Ben Arfi[3] · Rabeh Khalfaoui[4]

Abstract

This study was aimed at investigating the determinants of environmental sustainability in 86 countries from 2007 to 2018. The natural gradient boosting (NGBoost) algorithm was implemented along with five machine learning models to forecast the trends of CO_2 emissions. In addition, the SHapley Additive exPlanation (SHAP) technique was used to interpret the findings and analyze the contribution of the individual factors. The empirical results indicated that the predictions obtained using NGBoost were more accurate than those obtained using other models. The SHAP value exhibited a positive correlation among the amount of CO_2 emissions, economic growth, and opportunity entrepreneurship. A negative correlation was observed among the governance, personnel freedom, education, and pollution.

Keywords CO_2 · Gross domestic product · Institutional conditions · Environmental degradation · NGBoost · SHAP value

1 Introduction

According to independent analyses by the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA), 2019 was the second warmest year on record and the end of the warmest decade (2010–2019) since modern recordkeeping began in 1880 [1]. The last 5 years have been the warmest among the last 140 years [2]. Climate change is influencing all the countries worldwide. In particular, climate change can disrupt national economies and influence the lives of citizens. Moreover, with the changing weather conditions, sea levels are increasing, and weather events are becoming increasingly extreme. The main challenges among the 17 sustainable development goals (SDGs) were to address climate change and adopt environmental protection measures. In general, SDGs are a call to action for all nations with low, middle, and high incomes to promote prosperity while protecting the planet. As countries move toward rebuilding their economies after COVID-19, it is expected that the stimulus packages can shape the economy of the twenty-first century in a clean, green, healthy, secure, and more resilient manner. In fact, the recent crisis is an opportunity to implement deep and systemic changes to create a more sustainable economy that is beneficial to both people and the planet. To address the climate emergencies, post-pandemic stimulus plans must trigger long-term systemic changes that can alter the trajectory of CO_2 levels in the atmosphere.

In recent years, governments worldwide have spent considerable time and effort in developing plans to chart a safer and more sustainable future for the citizens. Incorporating these plans at present, as part of recovery planning, can help the world recover better from the current crisis. This aspect is a strong motivation to review the energy policy of developed countries from the perspective of climate change. Governments have increasingly sought to strengthen environmental regulations or laws and generate environmental awareness. In this regard, this study is aimed at examining the impact of both these policy instruments on CO_2 emissions.

✉ Wissal Ben Arfi
wbenarfi@edcparis.edu

1 Institute of Sustainable Business and Organizations, Confluence: Sciences Et Humanités - UCLY, ESDES, 10 place des archives, 69002 Lyon, France

2 CEREFIGE Laboratory, University of Lorraine, Nancy, France

3 OCRE Research Lab, EDC Paris Business School, 74 - 80 rue Roque de Fillol, CS 10074, 92807 Puteaux Cedex, France

4 Laboratoire de recherche en Économie et Gestion: LR18ES27, Faculté des Sciences Economiques et de Gestion de Sfax, Sfax, Tunisia

Although the relationships among energy consumption, CO_2 emissions, and economic growth have been widely debated, the results have been inconclusive. In recent decades, the field of energy policy has undergone considerable development owing to the use of increasingly sophisticated research methods [3, 4]. Nevertheless, this study demonstrates the need for more rigors in quantitative energy and environment research, driven by recent developments. In particular, empirical methodologies in energy and CO_2 emission research involve notable limitations in terms of the validity of the estimated coefficients and their elasticities because the applied tests are not based on an appropriate quantitative framework. For example, in many analyses, diagnostic statistics and specification tests that are necessary to obtain unbiased and consistent regression results are often not performed or considered suitably.

The key contributions of this study are as follows. First, seven factors affecting CO_2 emissions are examined comprehensively. The factors are divided into three categories: economic environment, legislative environment, and environmental awareness. To the best of our knowledge, this is the first study in which the proposition of Akerlof [5] regarding the impact of democratic participation practices on CO_2 emissions is evaluated. Two indicators are considered to measure democratic participation: personal freedom and fair elections and political participation in governance. Second, this study examines alternative tools that can overcome the limitations of the existing quantitative approaches. Reliable alternative tools based on advanced machine learning models, namely, natural gradient boosting (NGBoost) techniques, XGBoost, LightGBM, neural networks, and random forest (RF) algorithms are considered. Notably, in the existing work, the gradient boosting framework has not been used to examine the relationship among CO_2 emissions, economic growth, energy use, and institutional conditions. Third, the importance of the input variables on environmental degradation is evaluated, as none of the existing empirical studies has employed the SHapley Additive exPlanation (SHAP) value and game theories to integrate interpretable machine learning in analyzing the complex nonlinear behavior of CO_2 emissions.

The remaining paper is structured as follows. Section 2 discusses the existing literature. Section 3 presents the methodology and considered data. Section 4 describes the experimental results and discusses the main findings. Section 5 presents the concluding remarks and describes the scope for future work.

2 Literature Review

Researchers have explained the high intensity of carbon emissions considering the increase in energy use owing to the rapid economic growth [6]. Among carbon emission studies, the peak theory of pollutant emissions has been widely accepted in many countries [7, 8]. CO_2 emission peaks have been tested using the environmental Kuznets curve (EKC) [9, 10]. In particular, an inverted U-shaped relationship was considered to exist between the gross domestic product (GDP) and pollutant emissions, as the level of environmental damage first increases to a peak with an increasing GDP and later declines [11]. This observation is aligned with the peak of energy use, which corresponds to a regulatory trend in the long term. In the industrialization phase, energy use initially rises to a peak value and later declines. The turning point is caused by the structural changes from the higher energy consumption of heavy industries to the low consumption of light industries. Moreover, the product structure shifts from general value-added to high value-added products and from material production to knowledge production [12]. Many empirical studies have approved this U-shaped relationship. Ben Cheikh et al. applied the nonlinear panel smooth transition regression to examine the Middle East and North African (MENA) region and reported on the inverted U-shaped pattern for the impact of energy on CO_2; moreover, the authors reported that an increase in the GDP significantly affected the pollutant emissions, thereby increasing the energy consumption of heavy industries [13]. Ehigiamusoe et al. examined the influence of the energy use in the CO_2 emissions–income relationship for 64 middle-income countries and quantified the effect of the GDP on the CO_2 emissions at different levels of energy use [14]. The authors applied multiplicative interaction models and noted that energy consumption affected the relationship between CO_2 emissions and income. Thus, we hypothesize the possible effects suggested from the theory.

> Hypothesis (H) 1: CO_2 emissions follow a Kuznets curve which increases with economic development and then decreases.

The relationship between entrepreneurship and environmental quality has attracted considerable attention from policymakers in the search for solutions to attain environmental sustainability. According to Omri, a U-shaped relationship exists between entrepreneurship and pollution, albeit only in high-income countries [15]. Dhahri and Omri [16] concluded that entrepreneurship contributes positively to the economic and social aspects of sustainable development and negatively to the environmental dimension. Moreover, several authors [17, 18] investigated the effect of the entrepreneurship opportunity on the quality of the environment and indicated a negative relationship between the two variables. Hence, we assume that the level of entrepreneurship opportunity is associated positively with CO_2 emissions.

H2: The level of entrepreneurial opportunity is associated with an increase in CO_2 emissions.

In this regard, to achieve low-carbon objectives, although the overall trend of the CO_2 Kuznets curve cannot be adjusted, factors to adjust the shape of the EKC can be identified [19]. When executing a low-carbon environmental policy, the radius of the U-shaped curve is expected to be smaller, and a horizontal path can be generated in which the inflection point occurs earlier [20]. A key factor to flatten the curve pertains to environmental law and regulations. Considering the data of China, Yin et al. [21] concluded that the EKC under liberal environment regulations gradually tended to the horizontal orientation to fall below the per capita CO_2 emissions and required considerable time to reach the inflection point. In contrast, the EKCs under stringent environment regulations increased rapidly, and after reaching the inflection point, the pollutant emissions reduced rapidly [21]. Moreover, Pang et al. reported that the government policies on environmental regulations could enhance energy savings and emission reductions [22]. In this regard, effective governance is essential to boost the impact of environmental regulations. According to Samimi et al., environmental governance is necessary to protect environmental quality and ensure sustainable use of resources [23]. Different measures of governance exert different effects on the CO_2 emissions [24], and the dimensions of governance include the institutional quality, quality of service, corruption index, accountability civil rights, and political liberties [25, 26].

The daily actions of people form the natural and social environment in which they reside according to their needs and their consumption of resources (water, food, and energy), the waste they produce and the way in which they treat it, and the government policies they approve. Numerous studies have shown that the direct influence on the actions of individuals is derived from information and awareness [26–28]. Furthermore, the measures to mitigate climate change cannot be implemented without public encouragement and engagement. Environmental education is a key element in establishing the ecological quality of a nation [29]. It has been reported that an increase in environmental awareness can serve as a tool to formulate energy policies [30]. Akerlof explains the relations between knowledge, attitudes, and behavior. He suggests that in some cases, public awareness as an end goal is a highly necessary political and programmatic goal for governments, effectively an inevitable goal for liberal democracies. Akerlof recommended three conditional elements to effectively exploit environmental awareness to solve problems related to the environment. First, a policy must be established to facilitate behavioral adjustment among individuals. Second, to achieve satisfactory results, decision-making institutions must involve

democratic participation practices. Finally, there must be a direct focus for community-level changes in education, morals, and values [5]. Our final hypothesis captures these necessary factors of environmental awareness:

H3a: The level of education is associated with a decrease in CO_2 emissions.
H3b: The level of governance is associated with a decrease in CO_2 emissions.
H3c: The level of personal freedom is associated with a decrease in CO_2 emissions.

3 Methodology

3.1 Machine Learning Models

3.1.1 Linear Regression

Linear regression (LR) represents the gold standard technique in the machine learning domain. The method of ordinary least squares (OLS) is commonly used to estimate the parameters of the intercept and slope regression. The model can be expressed as follows:

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \varepsilon_i \tag{1}$$

where Y is the output variable, X_i is the explanatory variable, and β_i is the parameter estimated through OLS regression. The limitation of LR is that it only clarifies the interaction between the mean input variables and output variables. Nevertheless, the functional form of the link between the target variable Y and predictor X can be obtained in a flexible manner through advanced machine learning models.

3.1.2 Neural Network Regression

In the neural network (NN) regression technique, a forward-structured artificial neural network is used to map a collection of input vectors to the output vector collection. The NN consists of several layers of nodes, with each layer linked to the subsequent layer [31]. In addition to the input nodes, each node is a neuron with a nonlinear activation function that can be expressed as follows:

$$Y_j = \frac{1}{1 + e^{-\sum_{i=1}^{n} W_{ij} X_i + \alpha_j}}, \tag{2}$$

where Y_j is the predicted value of the jth neuron, X_i is the input variable of the ith neuron, and W_{ij} is the weight between the ith and jth neurons in the following layer.

In the literature, many activation functions have been used, such as tangent hyperbolic, sigmoid, sinusoid,

logistic, and wavelet. They are represented by the following equations:

Tangent hyperbolic:

$$\psi_1(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \tag{3}$$

The sigmoid function, also called the logistic function:

$$\psi_2(x) = \frac{1}{1 + e^{-x}}, \tag{4}$$

Morlet wavelet function:

$$\psi_3(x) = e^{-b_1 x^2}(\cos(b_2 x)), \tag{5}$$

where b1 and b2 are actual variables that will be adjusted as weights between neurons during training. We used the logistic function in this paper since it better reflects the asymmetric behavior of the economic cycle. Furthermore, this function highlights that the expansion and recession phases have different dynamics. Usually, backpropagation algorithms are used to run multi-layer perceptron NNs, which represent the most commonly deployed NN. In the basic framework, gradient search techniques are used to minimize the mean square error between the real output value of the network and the predicted output value [32].

3.1.3 RF Algorithm

In recent years, RFs have attracted considerable attention in the social science domain owing to their high efficiency, ease of execution, and low computational costs. This ensemble learning methodology was developed by Breiman, centered on creating multiple regression trees [33]. The bootstrap samples obtained from each tree were used to train the whole training set [34]. The output can be obtained as follows:

$$\hat{Y} = \frac{1}{q}\sum_{i=1}^{q} h_k(X) \tag{6}$$

where $h(x)$ is a set of kth learner random tree and X is the vector of the input variables. A new training set is generated by replacing the original data for each constructed regression tree. To enhance the predictive capacity of the model, the model hyperparameters must be tuned using the validation dataset.

3.1.4 LightGBM Algorithm

LightGBM, proposed by Ke et al. [35], is a robust gradient boosting framework, which is less computationally expensive and has a higher training speed and lower memory

consumption compared to those of the gradient boosting decision tree algorithms. The LightGBM algorithm is based on two innovative strategies, i.e., gradient-based one-side sampling and unique function bundling. The final output model can be established as follows:

$$\hat{Y} = \sum_{t=0}^{M-1} f_t(X_i), \tag{7}$$

where $f_t(X_i)$ is the learned function of the tth decision tree and M is the number of trees. Details regarding the calculation procedures of LightGBM can be found elsewhere [35]. The predictive performance is considerably affected by the hyperparameters, and thus, before using LightGBM, the number of hyperparameters must be determined.

3.1.5 NGBoost Algorithm

Typically, gradient boosting approaches can obtain the most precise predictions over standardized or tabular input data. Recently, an innovative approach named NGBoost was developed [36], which could be used to calculate the statistical uncertainty through the gradient boosting technique by using probabilistic forecasts (including actual valued outputs). In general, NGBoost requires considerably less expertise to operate than the other relevant approaches and performs well on standard metrics. NGBoost achieves notably satisfactory results on small data sets. The predicted output can be expressed as follows [36]:

$$\hat{Y} = \theta^{(0)} - \gamma \sum_{m=1}^{M} \rho^{(m)} g^{(m)}(X), \tag{8}$$

where θ is obtained by a mixture of additive M base learner outputs, $g^{(m)}$ is the set of base learners, $\rho^{(m)}$ is the scaling factor per stage, and γ is the common learning rate.

Although the performance of the NGBoost is comparable to that of the existing probabilistic regression approaches, it exhibits several other advantages. Specifically, NGBoost is modular, scalable, and easy to use. The algorithm can be scaled to handle large variables or observations with reasonable complexity, in contrast to classical boosting algorithms.

3.1.6 XGBoost Algorithm

XGBoost [37] is a profiling technique widely used by the scientific community to solve classification and regression problems. The algorithm combines strong classifiers to obtain a strong tree model using a serial training process [38]. Overfitting is prevented by introducing a regularization term. New trees are constantly generated and trained to fit the last iteration. The predicted value can be defined as follows:

$$\hat{Y} = \sum_{k=1}^{K} f_k(X_i), \tag{9}$$

where K denotes the number of trees and $f_k(X_i)$ is the newly generated tree model of the input X_i for the kth tree.

When designing the XGBoost with enhanced predictive performance, suitable hyperparameters must be calculated. Specifically, an appropriate set of values for the hyperparameters essential to the establishment of the model should be defined and tested. In the present study, grid search with cross-validation was employed to identify the optimal parameters.

3.1.7 Parameter Optimization

In the grid search technique, the hyperparameters are tuned to evaluate the optimum values for a given model. This aspect is important as the output of the entire model depends on the defined hyperparameters. In general, machine learning algorithms involve several hyperparameters to be tuned. Adequate hyperparameter regulation can help avoid model overfitting [39].

In this study, a grid search with 10 cross-validations (GridSearchCV) was performed to tune the RF, NN, Light-GBM, NGBoost, and XGBoost. Appendix Table 6 lists the optimal parameters considered in this study.

3.2 Shapley Additive Explanation

Usually, interpreting the results of machine learning models is very difficult owing to their complex "black box" architecture. The ability to correctly interpret the output of a prediction model is extremely valuable. It creates the appropriate user trust, provides information on improving a model, and helps to understand the modeled process. In response, several tools and methods have recently been suggested for interpreting the predictions of complex models, but it is often difficult to know how these methods relate and when one method is preferable to another. To solve this problem, Lundberg and Lee [40] presented a unified framework for the interpretation of predictions, SHAP. The SHAP value is a better tool than others, such as the local interpretable model-agnostic explanations (LIME), the partial dependence plot (PDP), or ELI5 algorithms, based on three characteristics. First, regarding the "overall interpretability," the SHAP values indicate how much each predictor contributes, positively or negatively, to the target variable. Second, regarding the "local interpretability," each observation obtains its own set of SHAP values. Third, SHAP values can be calculated for any tree model, whereas other methods use linear regression or logistic regression models as alternate models.

To understand and evaluate the drivers to estimate the CO_2 emissions and to compare the importance of individual features, we calculated the SHAP values [41]. In general, the SHAP values test each combination of predictors to assess the effect of each predictor. Based on the game theory and conditional assumptions, the SHAP values are likely to be consistent across various trees [42]. To calculate the SHAP values, the output of a tree centered on a subset of functions, S, is defined as $g_x(S) = [E(g(x))]$, and the SHAP values are expressed as

$$\phi_{i,j} = \sum_{S \subseteq N\{i,j\}} \frac{|S|!(M-|S|-1)!}{M!} [g_x(S \cup \{i\} - g_x(S)], \tag{10}$$

Table 1 Variable definition

Variable	Source and definition
CO_2	Annual carbon dioxide emissions in tons per capita. Source: World Development Indicators database (2020)
GDP Capita (GDP)	Gross domestic product (GDP) per capita in constant 2010 U.S. dollars. Source: World Development Indicators database (2020)
GCF	Annual gross capital formation as a percentage of GDP dollars. Source: World Development Indicators database (2020)
Energy Use (EU)	Annual amount of the used energy measured in kg as oil equivalent per capita. Source: World Development Indicators database (2020)
Governance (GOU)	The governance pillar measures the countries' performance in four areas: effective and accountable government, fair elections and political participation, rule of law, and level of a country's democracy. Source: Legatum Institute database (2019)
Business Environment (BE)	The entrepreneurial opportunity indicates the entrepreneurial climate in which citizens can pursue new ideas and opportunities to improve their lives, leading to more wealth and higher social wellbeing. Source: Legatum Institute database (2019)
Education (EDU)	The education pillar measures the countries' performance in three broad areas: access to education, quality of education, and human capital. Source: Legatum Institute database (2019)
Personal Freedom (PF)	The personal freedom pillar measures the countries' performance in three areas: basic legal rights, individual freedom, and social tolerance. Source: Legatum Institute database (2019)

Table 2 Descriptive statistics

Variable	Min	Max	Nb.obs	Mean	Sd.dev	Skew	Kurtosis	Shapiro–Wilk stat	*p*-value
CO_2	0	51.9336	1032	4.840	6.134	2.919	13.255	0.725	0.000
EU	0	18,167.596	1032	1815.571	2453.826	2.546	9.915	0.731	0.000
GCF	−1.376	0.6738	1032	0.0399	0.132	−0.986	14.881	0.897	0.000
GDP	0	111,968.349	1032	18,694.432	22,639.793	1.622	2.268	0.768	0.000
BE	27.488	76.736	1032	53.462	10.385	0.209	−0.500	0.977	0.000
GOU	24.161	86.679	1032	53.483	15.830	0.5292	−0.937	0.927	0.000
EDU	18.710	87.462	1032	66.648	13.151	−0.908	0.734	0.939	0.000
PF	25.185	92.848	1032	62.142	17.062	−0.102	−1.018	0.968	0.000

where *M* is the number of input variables. To improve the interpretability of gradient boosting decision tree algorithms, a new explanation tool was proposed [41], which directly measured the local feature interaction effects. In particular, TreeExplainer, an explanation technique for trees allows the optimal local explanations to be computed considering the desirable properties according to the game theory. In this study, the SHAP explanation of the optimal model was examined through Python packages, to assess the importance of each factor.

4 Data and Variables

Panel data of 86 countries for the period 2007–2018 were considered. The world development indicators (WDI, 2020) were considered to extract the CO_2 emissions (metric tons per capita), gross capital formation (% to the total GDP), energy use (kg as oil equivalent per capita), and GDP per capita (constant 2010 U.S). For the institutional variables pertaining to governance, education, personal freedom, and business environment, the Legatum Institute database (2019) was considered. The definitions of the considered variables are presented in Table 1. The causal link among the CO_2, energy use, economic growth, and institutional conditions can be expressed as

$$CO2 = f(GDP, EU, GOU, EDU, PF, BE, GCF) \qquad (11)$$

Figure 1 shows the intensity of the relationships among the variables. The correlation matrix indicates that the pairs EU-CO_2, GDP-GOU, BE-GOU, and GOU-PF are strongly correlated.

Table 2 summarizes the statistics for the considered variables. The EU and GDP exhibit the largest standard deviation and mean for all the countries. The higher standard deviation also implies that the EU and GDP series is more volatile than the other variables. Moreover, the skewness coefficients for the GCF, EDU, and PF are negative and those for the CO_2, EU, GDP, BE, and GOU time series are positive.

Furthermore, the kurtosis coefficients for all the time series are positive, except for those of BE, GOU, and PF. None of the values for the kurtosis is equal to three, suggesting that the fat-tailed distribution occurs. The test for normality is indicated by the Shapiro–Wilk statistic, and a non-normal distribution is suggested.

5 Results and Discussion

5.1 Nonlinear Unit Root Tests

The initial stage of the analysis is to investigate the properties of the data. To experimentally evaluate whether the macroeconomic variables are stationary, five nonlinear unit root tests are applied. We used (FFFFF) [43], (CEO) [44], (OSHa) [45], (OY) [46], and (OEHa) [47] tests. The source of nonlinearity of the employed unit root tests has been identified as a time-dependent nonlinearity, as well as a
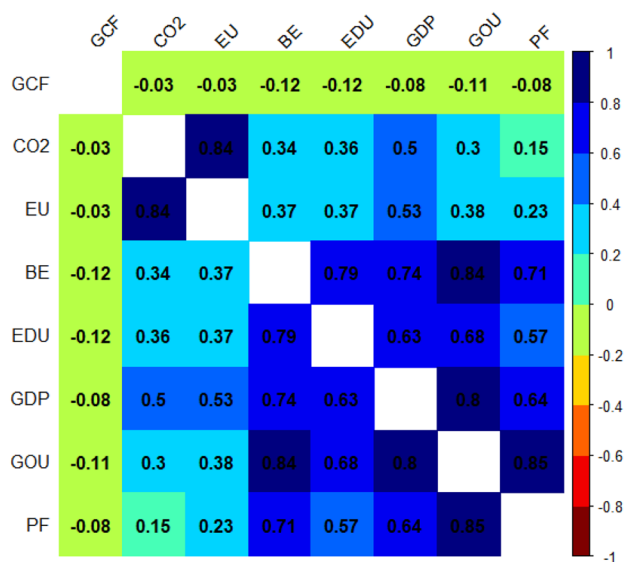
Fig. 1 Correlation matrix for the different variables

Table 3 Time-dependent nonlinearity analysis

Variables	FFFFF		CEO
	Intercept only	Intercept and time trend	
CO_2	−6.317*** (0.000)	−7.200*** (0.000)	−11.273*** (0.000)
ΔCO_2	−7.850*** (0.000)	−6.041*** (0.000)	−6.293*** (0.000)
EDU	−4.323*** (0.000)	−3.987*** (0.000)	−4.582*** (0.000)
ΔEDU	−6.292*** (0.000)	−12.932*** (0.000)	−12.926*** (0.000)
GDP	−7.097*** (0.000)	−5.936*** (0.000)	−6.569*** (0.000)
ΔGDP	−9.446*** (0.000)	−12.900*** (0.000)	−12.906*** (0.000)
GCF	−16.349*** (0.000)	−16.298*** (0.000)	−16.798*** (0.000)
ΔGCF	−14.949*** (0.000)	−14.95*** (0.000)	−14.957*** (0.000)
GOU	−5.225*** (0.000)	−3.629*** (0.000)	−5.330*** (0.000)
ΔGOU	−6.501*** (0.000)	−12.869*** (0.000)	−12.866*** (0.000)
EU	−5.848*** (0.000)	−6.831*** (0.000)	−6.538*** (0.000)
ΔEU	−8.048*** (0.000)	−6.612*** (0.000)	−11.209*** (0.000)
BE	−5.834*** (0.000)	−5.945*** (0.000)	−6.064*** (0.000)
ΔBE	−6.740*** (0.000)	−12.692*** (0.000)	−22.461*** (0.000)
PF	−4.862*** (0.000)	−4.894*** (0.000)	−5.294*** (0.000)
ΔPF	−6.269*** (0.000)	−12.740*** (0.000)	−12.729*** (0.000)

Statistics are *p*-values obtained: The parenthesis below the test. ** implies a significant coefficient at the 5% level, and *** implies a significant coefficient at the 1% level

combination of the time-dependent nonlinearity and a state-dependent nonlinearity, which may be referred to as a hybrid form of nonlinearity. The results of nonlinear unit root tests are reported in Tables 3 and 4.

Based on the FFFFF and CEO tests, we found that all variables are stationary at significance level and first difference. Moreover, the results tabulated in Table 3 suggested that all variables have a stationary process and nonlinear

trend around the deterministic component at various significance levels.

According to Omay et al. [48], these nonlinearities may emerge simultaneously in the data generating process to identify these hybrid structures (Table 4). OSHa unit root test provides evidence of stationarity for seven stationary variables, except GCF. Moreover, OEHa (2018) test states the same results, whereas OY (2014) reveals stationarity

Table 4 Hybrid nonlinearity analysis

Variables	OSHa	OEHa		OY
		Intercept only	Intercept and time trend	
CO_2	−6.177*** (0.000)	−3.747*** (0.000)	−3.803*** (0.000)	−7.072*** (0.000)
ΔCO_2	−3.867*** (0.000)	−2.279*** (0.0229	−2.227*** (0.0262)	−12.304*** (0.000)
EDU	−5.039*** (0.000)	−3.383*** (0.000)	−3.120*** (0.001)	−3.325*** (0.000)
ΔEDU	−3.685*** (0.000)	−4.105*** (0.000)	−4.105*** (0.000)	−5.920*** (0.000)
GDP	−4.601*** (0.000)	−4.519*** (0.000)	−4.547*** (0.000)	−6.669*** (0.000)
ΔGDP	−3.721*** (0.000)	−3.812*** (0.000)	−3.812*** (0.000)	−6.865*** (0.000)
GCF	−1.248 (0.212)	−0.482 (0.63)	−0.489 (0.625)	−4.023*** (0.000)
ΔGCF	0.181 (0.856)	0.118 (0.906)	0.118 (0.906)	−2.312** (0.02)
GOU	−4.623*** (0.000)	−5.174*** (0.000)	−4.985*** (0.000)	−5.813*** (0.000)
ΔGOU	−3.618*** (0.000)	−3.930*** (0.000)	−3.928*** (0.000)	−8.515*** (0.000)
EU	−5.426*** (0.000)	−4.505*** (0.000)	−4.602*** (0.000)	−10.018*** (0.000)
ΔEU	5.777*** (0.000)	3.887*** (0.000)	3.872*** (0.000)	−2.062** (0.0396)
BE	−6.710*** (0.000)	−6.733*** (0.000)	−6.766*** (0.000)	−6.676*** (0.000)
ΔBE	−5.279*** (0.000)	−3.193*** (0.000)	−3.194*** (0.000)	−4.590*** (0.000)
PF	−5.918*** (0.000)	−6.068*** (0.000)	−3.256*** (0.000)	−6.578*** (0.000)
ΔPF	−2.997** (0.002)	−3.256*** (0.001)	−3.255*** (0.001)	−6.303*** (0.000)

Statistics are *p*-values obtained: The parenthesis below the test. ** implies a significant coefficient at the 5% level, and *** implies a significant coefficient at the 1% level

Table 5 Comparison of predicted CO_2 emissions

Model	RMSE	R^2	MAE	MSE
XGBoost	1.763	0.905	1.122	3.111
NGBoost	**1.755**	**0.907**	**1.108**	**3.083**
LightGBM	2.136	0.860	1.298	4.565
Neural networks	1.896	0.903	1.206	3.957
Random forest	1.850	0.895	1.146	3.425
Linear regression	2.069	0.885	1.328	4.282

Values in bold indicate the best predictive performance

for all variables. Our results are consistent with Omay et al. [48], who documented that future movements in macroeconomic variables may be predicted based on previous behavior, because time series data on the state level follow mean-reversing processes. To conclude, all the variables that will be included in machine learning models were found to be stationary.

5.2 Performance Evaluation of Machine Learning Models

We designed artificial intelligence models to predict CO_2 emissions. Six machine learning approaches, namely, the LR, RF, NN, LightGBM, NGBoost, and XGBoost, were applied. The training and test data were divided into a ratio of 70:30; 70% of the data were used for model training, and 30% were used to examine the efficiency and accuracy of the proposed model. Four metrics were considered to evaluate the performance of the forecasting models: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). As the predictor variables included several macroeconomic variables with varying sizes and domains, data normalization was performed to access a homogeneous

dataset. Specifically, we normalized the initial data series for all the variables (X_i).

Table 5 lists the R^2, MAE, RMSE, and MAPE values for the testing data for all the models. NGBoost exhibits the best predictive performance in terms of all the metrics, with considerably higher accuracy than that of the XGBoost, RF, LightGBM, NN, and LR. In addition, NGBoost exhibits the largest R^2 of 0.907 and smaller values of the MAE, RMSE, and MAPE compared to those of the five techniques. LR exhibits the worst performance. The NGBoost model considerably enhances the precision of the forecasting method. Moreover, because NGBoost is a complex nonlinear model, it provides the relative value of the individual features used to construct the model.

5.3 Feature Analysis

Using the measured SHAP function value, we compared the global importance of all the characteristics. The SHAP analysis was performed for different communities by using the trained NGBoost model. In particular, the SHAP values define the significance of a feature by evaluating the forecasted value with and without considering the feature.

Figure 2 presents the SHAP global feature importance plot for the optimal machine learning model. Seven variables notably influence the CO_2 emissions. Moreover, the highest contribution to pollution corresponds to the GDP, followed by energy use, education, personal freedom, governance, business environment, and gross capital formation. GDP appears to be the most important variable. A review of the relevant studies indicates that economic development (GDP per capita) is likely the most significant factor [40–43]. Furthermore, energy use appears as the second most important factor. These findings are in line with those reported previously [44–47]. However, energy consumption and economic

**Fig. 2** Contribution of input variables to the amount of CO_2 emissions
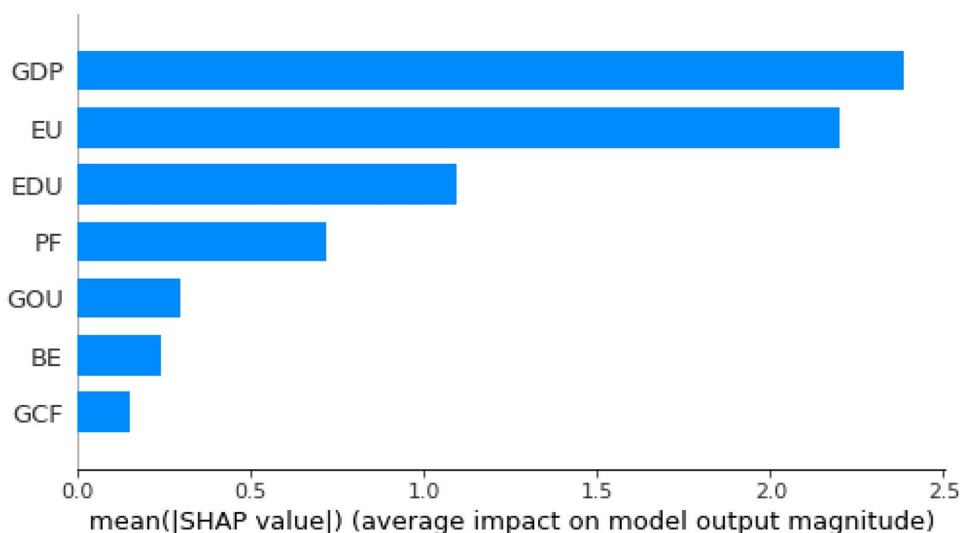
growth negatively influence the environment. Policymakers need to be very careful when choosing between taking care of the environment and stimulating economic growth. However, it is difficult to formulate economic and environmental policies while promoting economic growth and protecting the environment from degradation. This is why environmental awareness is a key factor in moderating the relationship between economic growth and environmental degradation. In fact, environmental awareness promotes environmentally friendly behavior without negatively affecting economic growth.

In Fig. 3, each point on the overview plot corresponds to the SHAP value for a characteristic of the feature. The location on the x-axis corresponds to the SHAP value, and the color indicates the intensity, ranging from low (blue) to high (red). Figure 3 indicates that strong positive relationships exist between the GDP and CO_2 emissions. In addition, higher values of the GDP result in higher values of the SHAP, corresponding to a higher probability of an increase in the CO_2 emissions. This observation is consistent with the previous findings [51], which indicated that the economic growth positively and significantly influences the CO_2 emission levels based on unbalanced panel data for 128 nations for the period 1990–2014. Energy use is the second most important variable; higher values of this factor are expected to aggravate environmental degradation. This finding is in line with those reported previously [52], which indicated that the CO_2 emissions increase owing to the increase in energy consumption. Education is the third most important variable, and it can be noted that higher values of education correspond to lower values of CO_2 emissions. This evidence is consistent with H3a that education enhances environmental awareness by understanding the harmful consequences of emissions and fosters environmentally friendly behavior.

The findings are in line with the prior work of Stef and Ben Jabeur [53]. Moreover, H3c is confirmed; high values of personal freedom decrease the volume of CO_2 emissions as stated by Brulle (2010, p. 91): "It is well known that political mobilization campaigns are more effective and legitimate if they engage citizens in a sustained dialog rather than treating them as a mass opinion to be manipulated. The importance of public participation in developing decisions that include concern about the natural environment has been stressed by numerous authors." Fig. 2 indicates that effective governance can moderate the amount of carbon emissions. These findings are in line with those reported previously [54] and those reported by Stef and Ben Jabeur, indicating that effective governance can mitigate environmental degradation and enhance environmental quality. Figure 3 shows the impact of the business environment on environmental degradation. In particular, although entrepreneurial opportunity increases the profits, the ecosystem is destroyed as the environmental quality is considered to be a luxury at this point. These findings confirm our second hypothesis (H2) that the level of entrepreneurship opportunity affects the quality of the environment. Our results are in line with those reported previously [15]. A slight positive effect can be noted in terms of the gross capital formation, although high levels of the gross capital formation may lead to considerable degradation of the ecosystem.

The SHAP values can be used to obtain the partial dependence plot, which shows the marginal effect of one or two variables on the predicted CO_2 emissions. Figure 4 shows the SHAP dependence plots of the NGBoost model, and the findings validate the aforementioned observations regarding the factors and outcome. Specifically, Fig. 4a shows that a nonlinear relationship does not exist between the GDP and CO_2, and initially, CO_2 volumes increase with

Fig. 3 SHAP summary plots for the CO_2 emissions

**(a)**

(b)

(c)

(d)

(e)
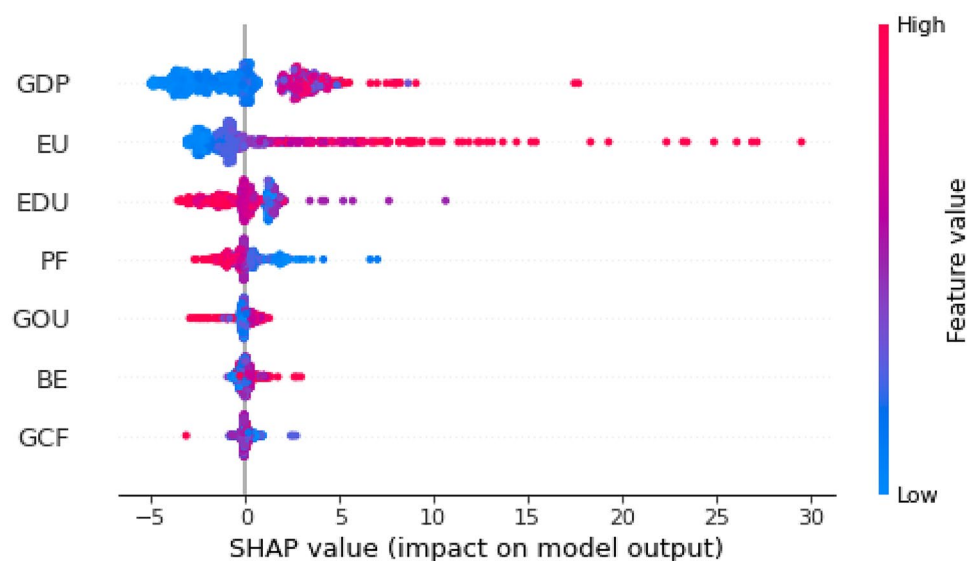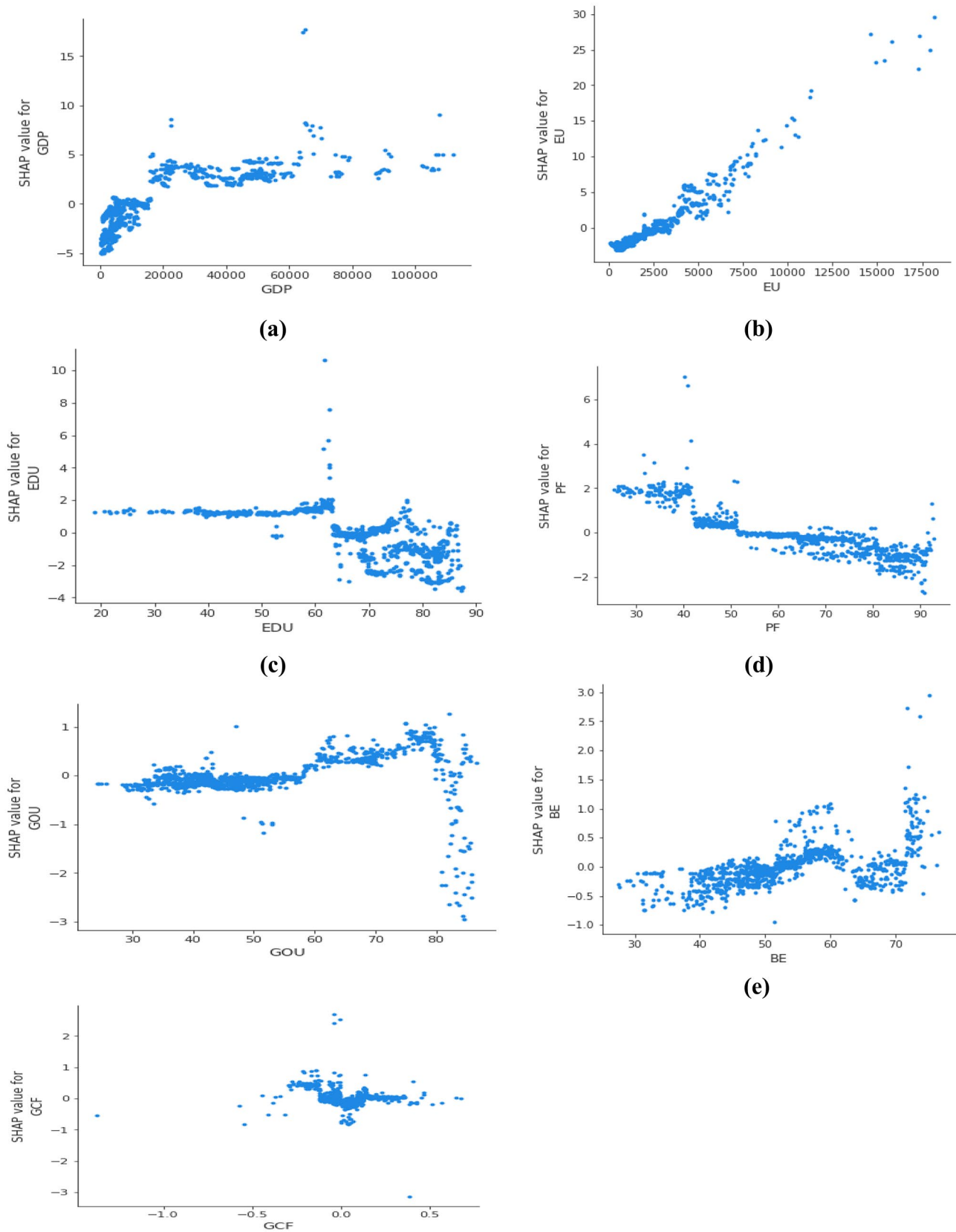
Fig. 4 SHAP dependency analysis performed using NGBoost. Dependency plots of **a** GDP, **b** EU, **c** EDU, **d** PF, **e** GOU, **f** BE, **g** GCF

increasing values of the GDP and then it decreases. The findings appear to confirm H1, the EKC hypothesis, according to which, the economic growth and environmental degradation exhibit an inverted U-shape. This result is in line with several existing reports pertaining to the EKC [49, 50]. Moreover, the dependence plot indicated the presence of a positive relationship between energy use and CO_2 emission volumes.

Using the SHAP interaction values, the effects of a feature on a single sample can be decomposed into the significant influence and interactions with other variables. Lundberg et al. proposed a modern method of interpretation in which the local factor interaction effects were calculated [41]. To perform the local interpretation, we selected two highly polluted countries, namely, Pakistan and Bahrain, and two highly clean countries, namely, Switzerland and France. The prediction of the output using the NGBoost model is illustrated in Fig. 5. The red and blue variables indicate that the outcome is higher and lower than the base value, respectively. Figure 5a shows the contribution of each feature to the model prediction for Pakistan in 2016. The base value is equal to 5.57, and the model prediction is 3.18. On the right side of the forecast value, the diagram reflects certain characteristics in blue that drive the prediction to lower values, and on the left side, in red, one can observe the characteristics that drive the prediction to higher values. It can be

noted that the GDP (1159) and EU (199) are associated with the decrease in CO_2 emissions. In contrast, the GOU, EDU, and PF are not notably associated with the increase in carbon dioxide emissions. In other words, emerging countries do not focus considerably on institutional conditions and climate change policies. In the case of Bahrain, as shown in Fig. 5b, the high values of EDU and EU cause the CO_2 emissions to decrease, and the GOU and GDP values cause the predicted output to increase. In other words, Arab countries involve less democratic aspects and personal freedom. Therefore, Arab countries as well as emerging authorities which have less democratic aspects and less personal freedom are recommended to target awareness raising as a main element of their communication campaigns and education programs in the pursuit of pro-environmental and practical decisions by the public. Figure 5c shows the local interpretation for Switzerland. The EDU, PF, GOU, BE, and EU are associated with lower CO_2 emissions. In other words, developed countries with a high degree of freedom, more educated people, and strong environmental laws tend to be more respected [53]. Nevertheless, the GDP contributes to environmental degradation. In the case of France, as shown in Fig. 5d, the EDU, PF, BE, and GCF are associated with lower values of CO_2 emissions. In contrast, the GDP is associated with higher CO_2 emissions.

(a)

(b)

(c)

(d)
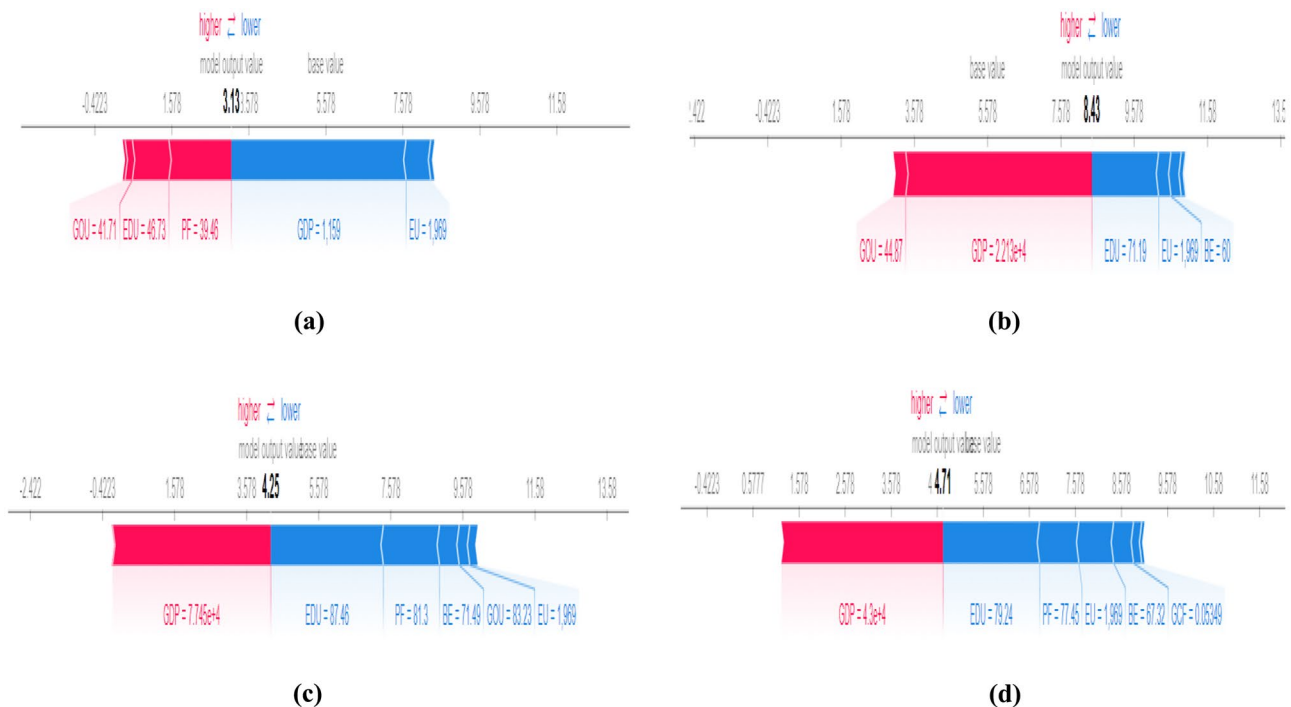
Fig. 5 SHAP force plots for the local interpretability, generated using NGBoost. Local interpretability (in 2016) for **a** Pakistan, **b** Bahrain, **c** Switzerland, **d** France. Feature values were listed at the bottom of the plot. Each group of features was ranked from center to both ends by the magnitude of their impact. Features that push the prediction higher (to the right) are shown in red, and those pushing the prediction lower are in blue

6 Conclusions

The causal relationships between the CO_2 emissions, economic growth, energy use, and institutional conditions were examined through machine learning models, which have attracted considerable interest in academic and policy debates. To interpret the complex and black-box models, such as those pertaining to the feature importance and local effects, the SHAP value was employed.

Environmental pollution is a multi-faceted challenge; every stage of environmental pollution is related to economic growth. Our empirical results show an EKC relationship between output growth and environmental degradation, which can be explained in a multidimensional way. The EKC literature emphasizes that there is no single policy that can reduce pollution levels and increase economic growth at the same time. Therefore, our study provides evidence that environmental awareness is a crucial process for reducing CO_2 emissions without negatively affecting economic growth. This observation is a strong reason for its consideration in public policies. Environmental awareness must be considered in public education and participatory decision-making processes for governments to maintain democracy. The methods by which environmental awareness is most likely to be achieved are not through campaigns using small messages; instead, they are more likely to be group processes within and between universities, schools, cities, communities, and nations [5]. Governments will still need to limit certain behaviors that damage the environment and encourage others. When implementing soft policy processes, environmental awareness should be the main component, but for optimal results several other behavioral components and interventions as well as regulations, taxes, and other financial incentives, must be implemented. In addition, further improvement in governance in emerging countries can help further reduce CO_2 emissions, as better governance implies better access to knowledge and democratic independence, thereby improving the awareness of citizens on environmental protection and fostering public interest and support for environmental legislation.

From a practical viewpoint, using advanced and interpretable machine learning models, the level of CO_2 emissions for large panel data can be predicted, thereby enabling the interpretation of the effect of each feature. Owing to this interpretability, machine learning is an effective tool for organizations and policymakers who require a deep understanding of the data and to help energy-consuming countries reduce excessive energy consumption to improve the environmental quality.

Nevertheless, this study involves two main limitations. Future work can be aimed at extending the proposed research framework by integrating more institutional conditions factors such as food security, health, and poverty. Second, more advanced machine learning and interpretable frameworks must be incorporated. As future directions, we suggest the following. (i) For a predictive accuracy purpose, hybrid predictive approaches can be developed, for instance, by combining the complex network and traditional artificial neural network (ANN) for carbon emissions forecasting. (ii) Other types of machine learning techniques can be used, such as support vector machine (SVM) and deep learning models: convolutional neural network (CNN) and long short-term memory (LSTM). (iii) Future research studies may also focus on the key sectors of different economies, such as energy, industrial, and services.

Author Contribution Sami Ben Jabeur: conceptualization, methodology, writing, software & data curation. Houssein Ballouk: conceptualization, methodology, formal analysis, writing–original draft. Wissal Ben Arfi: resources, conceptualization, review & editing. Rabeh Khalfaoui: supervision, writing–review & visualization.

Declarations

Appendix

Table 6 Hyperparameter optimization based on GridSearchCV

Model	Optimal parameters
NGBoost	• Base: DecisionTreeRegressor • criterion = friedman_mse • max_depth = 4 • minibatch_frac: 1.0
XGBoost	• colsample_bytree: 0.7 • learning_rate: 0.03 • max_depth: 7 • min_child_weight: 4 • n_estimators: 500 • nthread: 4 • objective: reg:linear • silent: 1 • subsample′: 0.7
LightGBM	• max_depth: 7 • num_leaves: 50
NN	• activation: ReLU • alpha′: 0.0001 • hidden_layer_sizes: (100, 50) • learning_rate: adaptive • solver: adam
RF	• max_depth: 110 • max_features:3 • min_samples_leaf: 3 • min_samples_split: 8 • n_estimators: 300

References

1. Brown, K. (2020, January 15). NASA, NOAA Analyses Reveal 2019 Second Warmest Year on Record. *NASA*. Text. Retrieved October 30, 2020, from http://www.nasa.gov/press-release/nasa-noaa-analyses-reveal-2019-second-warmest-year-on-record

2. Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). NASA Goddard Institute for Space Studies. *Journal of Geophysical Research: Atmospheres, 124*(12), 6307–6326. https://doi.org/10.1029/2018JD029522

3. Wu, L., Liu, S., Liu, D., Fang, Z., & Xu, H. (2015). Modelling and forecasting CO2 emissions in the BRICS (Brazil, Russia, India, China, and South Africa) countries using a novel multi-variable grey model. *Energy, 79*, 489–495. https://doi.org/10.1016/j.energy.2014.11.052

4. Pérez-Suárez, R., & López-Menéndez, A. J. (2015). Growing green? Forecasting CO2 emissions with environmental Kuznets curves and logistic growth models. *Environmental Science & Policy, 54*, 428–437. https://doi.org/10.1016/j.envsci.2015.07.015

5. Akerlof, K. (2017). When Should Environmental Awareness Be a Policy Goal? In V. Loreto, M. Haklay, A. Hotho, V. D. P. Servedio, G. Stumme, J. Theunis, & F. Tria (Eds.), *Participatory sensing, opinions and collective awareness* (pp. 305–336). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-25658-0_15

6. Ahmad, N., Du, L., Lu, J., Wang, J., Li, H.-Z., & Hashmi, M. Z. (2017). Modelling the CO2 emissions and economic growth in Croatia: Is there any environmental Kuznets curve? *Energy, 123*, 164–172. https://doi.org/10.1016/j.energy.2016.12.106

7. Le, H. P., & Ozturk, I. (2020). The impacts of globalization, financial development, government expenditures, and institutional quality on CO2 emissions in the presence of environmental Kuznets curve. *Environmental Science and Pollution Research, 27*(18), 22680–22697. https://doi.org/10.1007/s11356-020-08812-2

8. Sarkodie, S. A., & Ozturk, I. (2020). Investigating the environmental Kuznets curve hypothesis in Kenya: A multivariate analysis. *Renewable and Sustainable Energy Reviews, 117*, 109481. https://doi.org/10.1016/j.rser.2019.109481

9. Dong, K., Sun, R., Li, H., & Liao, H. (2018). Does natural gas consumption mitigate CO2 emissions: Testing the environmental Kuznets curve hypothesis for 14 Asia-Pacific countries. *Renewable and Sustainable Energy Reviews, 94*, 419–429. https://doi.org/10.1016/j.rser.2018.06.026

10. Kong, Y., & Khan, R. (2019). To examine environmental pollution by economic growth and their impact in an environmental Kuznets curve (EKC) among developed and developing countries. *PLoS One, 14*(3), e0209532. https://doi.org/10.1371/journal.pone.0209532

11. Grossman, G. M., & Krueger, A. B. (1995). Economic growth and the environment. *The Quarterly Journal of Economics, 110*(2), 353–377. https://doi.org/10.2307/2118443

12. Sun, J. W. (1999). The nature of CO2 emission Kuznets curve. *Energy Policy, 27*(12), 691–694. https://doi.org/10.1016/S0301-4215(99)00056-7

13. Ben Cheikh, N., Ben Zaied, Y., & Chevallier, J. (2021). On the nonlinear relationship between energy use and CO2 emissions within an EKC framework: Evidence from panel smooth transition regression in the MENA region. *Research in International Business and Finance, 55*, 101331. https://doi.org/10.1016/j.ribaf.2020.101331

14. Ehigiamusoe, K. U., Lean, H. H., & Smyth, R. (2020). The moderating role of energy consumption in the carbon emissions-income nexus in middle-income countries. *Applied Energy, 261*, 114215. https://doi.org/10.1016/j.apenergy.2019.114215

15. Omri, A. (2018). Entrepreneurship, sectoral outputs and environmental improvement: International evidence. *Technological Forecasting and Social Change, 128*, 46–55. https://doi.org/10.1016/j.techfore.2017.10.016

16. Dhahri, S., & Omri, A. (2018). Entrepreneurship contribution to the three pillars of sustainable development: What does the evidence really say? *World Development, 106*(2018), 64–77. https://doi.org/10.1016/j.worlddev.2018.01.008

17. He, J., Nazari, M., Zhang, Y., & Cai, N. (2020). Opportunity-based entrepreneurship and environmental quality of sustainable development: A resource and institutional perspective. *Journal of Cleaner Production, 256*, 120390. https://doi.org/10.1016/j.jclepro.2020.120390

18. Omri, A., & Afi, H. (2020). How can entrepreneurship and educational capital lead to environmental sustainability? *Structural Change and Economic Dynamics, 54*, 1–10. https://doi.org/10.1016/j.strueco.2020.03.007

19. Lin, B. Q., & Jiang, Z. J. (2009). Environmental Kuznets curve: the prediction and the analysis of influencing factors of the CO2 of China. *Manag. World*, pp. 27–36.

20. Munasinghe, M. (1995). Making economic growth more sustainable. *Ecological Economics, 15*(2), 121–124. https://doi.org/10.1016/0921-8009(95)00066-6

21. Yin, J., Zheng, M., & Chen, J. (2015). The effects of environmental regulation and technical progress on CO2 Kuznets curve: An evidence from China. *Energy Policy, 77*, 97–108. https://doi.org/10.1016/j.enpol.2014.11.008

22. Pang, R., Zheng, D., Shi, M., & Zhang, X. (2019). Pollute first, control later? Exploring the economic threshold of effective environmental regulation in China's context. *Journal of Environmental Management, 248*, 109275. https://doi.org/10.1016/j.jenvman.2019.109275

23. Samimi, A. J., Ahmadpour, M., & Ghaderi, S. (2012). Governance and environmental degradation in MENA region. *Procedia - Social and Behavioral Sciences, 62*, 503–507. https://doi.org/10.1016/j.sbspro.2012.09.082

24. Halkos, G. E., & Tzeremes, N. G. (2013). Carbon dioxide emissions and governance: A nonparametric analysis for the G-20. *Energy Economics, 40*, 110–118. https://doi.org/10.1016/j.eneco.2013.06.010

25. Dutt, K. (2009). Governance, institutions and the environment-income relationship: A cross-country study. *Environment, Development and Sustainability, 11*(4), 705–723. https://doi.org/10.1007/s10668-007-9138-8

26. Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology, 25*(3), 273–291. https://doi.org/10.1016/j.jenvp.2005.08.002

27. Jackson, T. (2014). Sustainable consumption. *Handbook of Sustainable Development*. Retrieved from https://www.elgaronline.com/view/edcoll/9781782544692/9781782544692.00029.xml

28. McKenzie-Mohr, D. (2011). *Fostering sustainable behavior: An introduction to community-based social marketing (Third Edition)*. New Society Publishers.

29. Zafar, M. W., Shahbaz, M., Sinha, A., Sengupta, T., & Qin, Q. (2020). How renewable energy consumption contribute to environmental quality? The role of education in OECD countries. *Journal of Cleaner Production, 268*, 122149. https://doi.org/10.1016/j.jclepro.2020.122149

30. Lillemo, S. C. (2014). Measuring the effect of procrastination and environmental awareness on households' energy-saving behaviours: An empirical approach. *Energy Policy, 66*, 249–256. https://doi.org/10.1016/j.enpol.2013.10.077

31. Salisu, A. A., Ebuh, G. U., & Usman, N. (2020). Revisiting oil-stock nexus during COVID-19 pandemic: Some preliminary results. *International Review of Economics & Finance, 69*, 280–294. https://doi.org/10.1016/j.iref.2020.06.023

32. Dou, M., Qin, C., Li, G., & Wang, C. (2020). Research on calculation method of free flow discharge based on artificial neural network and regression analysis. *Flow Measurement and Instrumentation, 72*, 101707. https://doi.org/10.1016/j.flowmeasinst.2020.101707

33. Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

34. Babar, B., Luppino, L. T., Boström, T., & Anfinsen, S. N. (2020). Random forest regression for improved mapping of solar irradiance at high latitudes. *Solar Energy, 198*, 81–92. https://doi.org/10.1016/j.solener.2020.01.034

35. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 30). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf

36. Duan, T., Anand, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A., & Schuler, A. (2020). NGBoost: Natural gradient boosting for probabilistic prediction. In H. D. III & A. Singh (Eds.), *Proceedings of the 37th International Conference on Machine Learning* (Vol. 119, pp. 2690–2700). PMLR. Retrieved from http://proceedings.mlr.press/v119/duan20a.html

37. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

38. Mo, H., Sun, H., Liu, J., & Wei, S. (2019). Developing window behavior models for residential buildings using XGBoost algorithm. *Energy and Buildings, 205*, 109564. https://doi.org/10.1016/j.enbuild.2019.109564

39. Rtayli, N., & Enneya, N. (2020). Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. *Journal of Information Security and Applications, 55*, 102596. https://doi.org/10.1016/j.jisa.2020.102596

40. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions (pp. 4768–4777). Presented at the Proceedings of the 31st international conference on neural information processing systems.

41. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

42. Lim, S., & Chi, S. (2019). Xgboost application on bridge management systems for proactive damage estimation. *Advanced Engineering Informatics, 41*, 100922. https://doi.org/10.1016/j.aei.2019.100922

43. Omay, T. (2015). Fractional frequency flexible Fourier form to approximate smooth breaks in unit root testing. *Economics Letters, 134*, 123–126. https://doi.org/10.1016/j.econlet.2015.07.010

44. Çorakcı, A., Emirmahmutoglu, F., & Omay, T. (2017). Re-examining the real interest rate parity hypothesis (RIPH) using panel unit root tests with asymmetry and cross-section dependence. *Empirica, 44*(1), 91–120. https://doi.org/10.1007/s10663-015-9312-4

45. Omay, T., Shahbaz, M., & Hasanov, M. (2020). Testing PPP hypothesis under temporary structural breaks and asymmetric dynamic adjustments. *Applied Economics, 52*(32), 3479–3497. https://doi.org/10.1080/00036846.2020.1713293

46. Omay, T., & Yildirim, D. (2014). Nonlinearity and smooth breaks in unit root testing. *Econometrics Letters, 1*(1), 2–9.

47. Omay, T., Emirmahmutoglu, F., & Hasanov, M. (2018). Structural break, nonlinearity and asymmetry: A re-examination of PPP proposition. *Applied Economics, 50*(12), 1289–1308. https://doi.org/10.1080/00036846.2017.1361005

48. Omay, T., Ozcan, B., & Shahbaz, M. (2020). Testing the hysteresis effect in the US state-level unemployment series. *Journal of Applied Economics, 23*(1), 329–348. https://doi.org/10.1080/15140326.2020.1759865

49. Omri, A. (2013). CO2 emissions, energy consumption and economic growth nexus in MENA countries: Evidence from simultaneous equations models. *Energy Economics, 40*, 657–664. https://doi.org/10.1016/j.eneco.2013.09.003

50. Shahbaz, M., Zakaria, M., Shahzad, S. J. H., & Mahalik, M. K. (2018). The energy consumption and economic growth nexus in top ten energy-consuming countries: Fresh evidence from using the quantile-on-quantile approach. *Energy Economics, 71*, 282–301. https://doi.org/10.1016/j.eneco.2018.02.023

51. Dong, K., Hochman, G., Zhang, Y., Sun, R., Li, H., & Liao, H. (2018). CO2 emissions, economic and population growth, and renewable energy: Empirical evidence across regions. *Energy Economics, 75*, 180–192. https://doi.org/10.1016/j.eneco.2018.08.017

52. Muhammad, B. (2019). Energy consumption, CO2 emissions and economic growth in developed, emerging and Middle East and North Africa countries. *Energy, 179*, 232–245. https://doi.org/10.1016/j.energy.2019.03.126

53. Stef, N., & Ben Jabeur, S. (2020). Climate change legislations and environmental degradation. *Environmental and Resource Economics*. https://doi.org/10.1007/s10640-020-00520-2

54. Omri, A., & Bel Hadj, T. (2020). Foreign investment and air pollution: Do good governance and technological innovation matter? *Environmental Research, 185*, 109469. https://doi.org/10.1016/j.envres.2020.109469