Energy Strategy Reviews 49 (2023) 101159

Contents lists available at ScienceDirect

Energy Strategy Reviews

journal homepage: www.elsevier.com/locate/esr

Quantifying the impact of energy consumption sources on GHG emissions in major economies: A machine learning approach

Mutaz AlShafeey [a,*], Omar Rashdan [b]

[a] *Institute of Data Analytics and Information Systems, Corvinus University of Budapest, Budapest, Fővám tér 13-15, H-1093, Hungary*
[b] *Faculty of Pharmacy, Middle East University, Amman, Airport Rd., 11831, Jordan*

ARTICLE INFO

Handling Editor: Mark Howells

Keywords:
Greenhouse gas emissions
Energy consumption sources
Machine learning
United States
China
European Union

ABSTRACT

This article aims to quantify the impact of different energy consumption sources on greenhouse gas (GHG) emissions for three major economies: the United States of America (USA), China, and the European Union (EU). To achieve this, energy consumption and GHG emissions data were obtained from "Our World in Data" for the period 1965–2021. Then, two machine learning techniques were utilized. Gradient Boosting (GB) was used to identify the major energy consumption sources contributing to GHG. While Artificial Neural Network (ANN) was used to quantify the effects of these major energy consumption sources on GHG emissions. The findings have significant implications for policymakers, as they suggest that effective strategies to reduce GHG emissions must be tailored based on the energy utilization sources of each country. Specifically, for the USA it was found that reducing coal consumption could be the most effective strategy to reduce GHG emissions, as increasing coal consumption by 25% would result in a 13% increase in GHG emissions. In contrast, increasing nuclear consumption by 25% in China would result in an 11% decrease in GHG emissions due to the displacement of fossil fuel-based energy sources. Increasing wind energy consumption by 25% in China would result in a 3% decrease in GHG emissions. In the EU, the study found that increasing oil consumption has a minor effect on GHG emissions while increasing coal consumption by 25% would result in an 11% increase in GHG emissions, highlighting the importance of reducing coal consumption. This study's originality lies in the use of machine learning techniques to identify the key energy consumption sources driving GHG emissions in the three major economies, as well as its specific recommendations for reducing emissions.

Credit author statement

Mutaz AlShafeey: Conceptualization, Methodology, Software, Data curation, Writing – original draft. Omar Rashdan: Methodology, Investigation, Validation, Writing- Reviewing and Editing.

1. Introduction

Greenhouse gases (GHG) are well known for their harmful impacts on our planet [1]; Carbon dioxide contributes to rising global temperatures, altered precipitation patterns, and sea level rise. Methane exacerbates climate change and ground-level ozone formation. While nitrous oxide has a high warming potential and poses risks to ecosystems and human health. The growing rates of GHG emissions have imposed substantial risks to human life and the overall environment [2]. Starting from the first industrial revolution, a gradual, but significant increase in average annual temperature that is associated with extreme weather events and severe temperatures was observed. This change has directly impacted agriculture as well as some other sectors [3–5]. The issue was exacerbated as the effects of climate change have become more serious with an average increase in global temperature between 0.5 and 1 °C over the past decade [5]. One of the reasons for this temperature rise is the high levels of GHG emissions associated with manufacturing and economic activities [3,6]. The United States (US), China, and the European Union (EU) are together responsible for more than half of the world's total GHG emissions; Accounting for 28% of the global emissions, China is the world's largest emitter of GHGs, followed by the US as contributing for 15% of global GHG emissions, while the EU contributed to 9% of the global emissions [7]. For these economies, energy consumption is one of the most contributors to GHG emissions, as over two-thirds of GHG emissions related to human activities are produced by energy sources highlighting the need for cleaner and more sustainable

* Corresponding author. Budapest, Fővám tér 13-15, H-1093, Hungary.
E-mail addresses: Mutaz.AlShafeey@uni-corvinus.hu (M. AlShafeey), orashdan@meu.edu.jo (O. Rashdan).

https://doi.org/10.1016/j.esr.2023.101159
Received 19 May 2023; Received in revised form 7 August 2023; Accepted 9 August 2023
Available online 16 August 2023

energy sources [8].

To reduce GHG emissions and mitigate the impact of climate change, low-carbon transition solutions can be beneficial [9]. In the energy sector, promoting such a transition requires either reducing energy consumption, increasing the utilization of low-carbon energy sources, or improving energy efficiency [10]. Reducing energy usage is a hard and complicated task as the growth in energy demand is normally a result of economic and population growth [11]. Yet, decoupling energy from economic growth is a hot topic that is gaining attention [12]. Utilizing low-carbon energy on the other hand might offer a more feasible solution [13]. Driven by cultivating renewable technologies such as Photovoltaic (PV) and wind turbines, it is possible to produce green energy with a marginal carbon footprint. On the other hand, improving energy efficiency offers relatively marginal benefits [14]. Current solutions for reducing GHG emissions have largely focused on interventional solutions such as carbon pricing and renewable energy subsidies [15]. However, the effectiveness of these solutions can vary depending on the level of energy consumption diversity in different economies.

Different sources of energy consumption have varying impacts on GHG emissions [16]. Coal, oil, and gas are reported as the most widely used sources of energy worldwide, which consequently sets them as the highest global contributors to GHG emissions [16]. Although Nuclear energy does not emit GHG emissions during operation which makes it seem like a low-carbon energy source [17], the life-cycle emissions of nuclear energy rely on high-carbon dependent factors such as uranium mining, fuel enrichment, and waste disposal [18]. That is in addition to the subsequent safety, waste management, and nuclear proliferation issues, which might make nuclear energy a less appealing option in certain scenarios [16]. Similarly, renewable energy sources, such as solar, wind, and hydro, have low GHG emissions, as they do not emit any GHGs during operation. Yet, their life-cycle emissions are still present and can vary widely depending on the production methods used [19].

Machine learning algorithms such as Artificial neural networks (ANN) and gradient boosting (GB) have shown superiority over traditional statistical methods in studying energy GHG emissions [20,21]. These algorithms have the ability to model complex nonlinear relationships, which is not reliable with traditional statistical methods [22]. Moreover, ANN and GB are highly adaptable and can learn from new data over time, making them useful for predicting future GHG emissions [23].

Even though the literature review on energy and GHG emissions covers a wide range of topics such as energy efficiency, renewable energy, carbon capture and storage, and climate change mitigation [24]; a noticeable scarcity of investigational research addressing the diversity of energy consumption sources' impact on GHG is apparent. Some of the previous studies focused on the relationship between one source of energy consumption such as PV, wind, hydropower, bioenergy, and GHG emissions as in Refs. [25,26]. Other studies only focus on one sector of energy consumption such as the transportation sector [27]. While some studies focus on energy consumption technologies like Carbon Capture and Storage (CCS) [28], energy storage [29], and smart grids [30,31]. Yet, there is a need for a comprehensive approach to address GHG emissions, including studying various energy consumption sources, practical solutions, and policy recommendations to reduce emissions. In this work, we investigate the energy consumption sources with the highest impact on GHG, tailored to each country/region, given the different utility landscapes of energy sources of the three major economies, i.e. the US, China, and the EU. This work contributes to the existing literature through its original estimation methodology along with its practical policy applications in the field of energy policy. This is done by developing region-specific machine learning predictive models to analyze the relationship between energy consumption and greenhouse gas (GHG) emissions. With the aim of providing region specific energy policy recommendations on the utility of energy sources with the highest reduction impact on GHG emissions.

2. Methods

To achieve the aims of this study, the method depicted in Fig. 1 was employed. First data was collected and processed as in section 2.1. Then, two machine learning techniques (ANN and Gradient Boosting) were utilized as described in section 2.2. Finally, based on the machine learning analyses the results and recommendations are derived as in sections 3 and 4.

The artificial neural network was implemented using Matlab R2023a, while gradient boosting was implemented using Python Jupiter environment. For both methods, the data was split into 80% for training and 20% for testing. Additionally, the data were subjected to 3 k-folds validation to ensure the robustness of the models.

2.1. Data collection and processing

The data collected and analyzed in this study pertains to energy consumption and greenhouse gas emissions of three major global players - the United States of America, China, and the European Union. The data collection process involved sourcing data from the "Our World in Data" database covering the period from 1965 to 2021. Specifically, the study focused on two key datasets: energy consumption by source and greenhouse gas emissions. The energy consumption by source dataset measured primary energy consumption in terawatt-hours (TWh). The dataset comprises nine different sources; i.e. geo biomass, biofuels, solar, wind, hydro, nuclear, gas, coal, and oil. On the other hand, the greenhouse gas emissions dataset comprises the emissions of carbon dioxide, methane, and nitrous oxide from all sources combined, in metric tons.

In this study, the collected data on energy consumption and greenhouse gas emissions by the USA, China, and EU were first normalized to eliminate the effects of different measurement units and scales on the analysis results. Without such, variables with larger ranges and magnitudes will dominate the analysis, while smaller variables may be ignored. Furthermore, normalization enhances the stability and robustness of machine learning models by avoiding issues such as gradient explosion and vanishing [32].

Regarding the validity of the data, "Our World in Data" calculates these emissions using data from Jones et al. [33] as per the Intergovernmental Panel on Climate Change (IPCC) methodology [34]. Recently, the IPCC has released an update to its methodology that improves

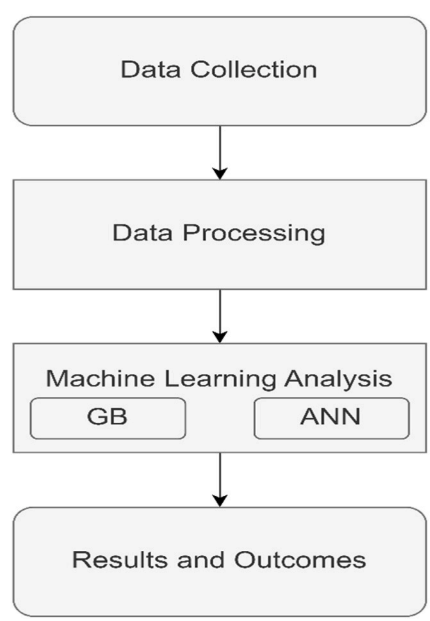

Fig. 1. General method flowchart.

transparency and reporting by ensuring that the methodology used to determine these inventories is based on the latest science. It also addresses gaps in the science that were identified, new technologies and production processes have emerged, or for sources and sinks that were not included in the earlier IPCC guidelines.

Note that the primary objective is to evaluate the impact of various energy sources on greenhouse gas (GHG) emissions. Hence, the study concentrates on the total annual emissions and energy consumption data of each source, which provides valuable insights into the relative contributions of different energy sources to GHG emissions, without necessitating a detailed utilization of the carbon life cycle of each individual source.

2.2. Machine learning models

In this study, two machine learning techniques, ANNs and GB were employed. The main objective of these techniques was to improve the accuracy of predictions and gain a deeper understanding of the factors that drive energy consumption and greenhouse gas emissions. We first trained and tested three separate ANN models for each region to forecast GHG emissions based on energy consumption data. Meanwhile, the Gradient Boosting method was utilized to identify the three most important sources among the nine collected energy consumption sources that affect GHG emissions. These most important sources were further investigated using the trained ANN models to simulate the effects of increasing their consumption from current levels (i.e. 2021) with 0.5% increments up to 25%, while controlling for all other sources, to isolate the target source impact for quantification.

2.2.1. Artificial neural network (ANN)

Feedforward ANN with Levenberg-Marquardt backpropagation algorithm was utilized as can be seen in Fig. 2.

During training, the input data is fed forward through the network using the following equations:

$$a(1) = x \tag{1}$$

$$a(l) = \sigma(h(l)) \tag{2}$$

$$a(l) = \sigma(w(l-1)a(l-1) + b(l-1)) \tag{3}$$

where x is the initial input data that is propagated forward through the network to produce the output. (l) is the activation vector of layer l, $h(l)$ is the weighted input vector of layer l, $w(l-1)$ is the weight matrix connecting layer $l-1$ to layer l, $b(l-1)$ is the bias vector of layer $l-1$, and σ is the activation function.

After the feedforward pass, the output of the network is compared to the target output, and the error is calculated. The weights and biases are then updated using the backpropagation with the Levenberg-Marquardt algorithm, which involves propagating the error back through the network and adjusting the weights and biases.

The delta rule is used to calculate the error at each layer:

$$\delta(L) = \nabla h(L)\mathscr{L} \odot \sigma'(h(L)) \tag{4}$$

$$\delta(l) = ((w(l))T\delta(l+1)) \odot \sigma'(h(l)) \tag{5}$$

where $\nabla h(L)\mathscr{L}$ is the gradient of the loss with respect to the output of the network, $\odot$ represents element-wise multiplication, and σ' is the derivative of the activation function. Note that $\delta(L)$ represents the error at the output layer, while $\delta(l)$ represents the error at layer l where l can range from 1 (the input layer) to $L-1$ (the layer before the output layer). The weights and biases are then updated using the following equations:

$$w^l := w^l - \Delta w^l \tag{6}$$

$$b^l := b^l - \Delta b^l \tag{7}$$

The weight and bias updates for layer l are then calculated using regularized Gauss-Newton equations to determine the tradeoff between the gradient descent and Gauss-Newton steps:

$$\Delta w^l = \left(J^T J + \lambda I\right)^{-1} J^T \delta^{l+1} \left(a^l\right)^T \tag{8}$$

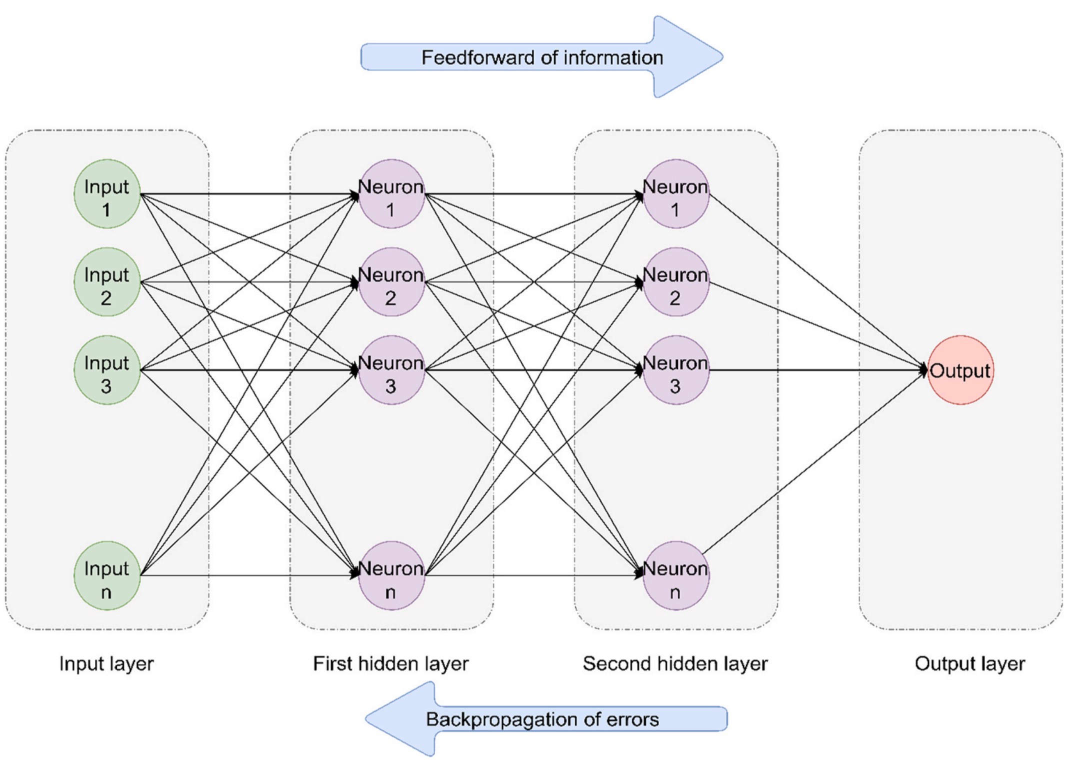

Fig. 2. A neural network with n inputs and one output.

$$\Delta b^l = \left(J^T J + \lambda I\right)^{-1} J^T \delta^{l+1} \tag{9}$$

where Δw^l and Δb^l are the weight and bias updates, respectively for layer l. λ is the Levenberg-Marquardt parameter, and J is the Jacobian matrix, which is defined as:

$$\mathrm{J}_{\mathrm{ij}}^{(\mathrm{l})} = \frac{\partial \mathrm{h}_{\mathrm{i}}^{(\mathrm{l})}}{\partial \mathrm{w}_{\mathrm{ij}}^{(\mathrm{l-1})}} \tag{10}$$

$$= \frac{\partial}{\partial w_{ij}^{(l-1)}} \left(\sum_k w_{ik}^{(l-1)} a_k^{(l-1)} + b_i^{(l-1)} \right) \tag{11}$$

$$= a_j^{(l-1)} \tag{12}$$

$$\delta^{(L)} = J^{(L)} \left(f\left(h^{(L)}\right) - y \right) \tag{13}$$

$$\delta^{(l)} = \left(J^{(l)}\right)^T \left(J^{(l)} \left(J^{(l)}\right)^T + \mu I \right)^{-1} \delta^{(l+1)} \tag{14}$$

$$\frac{\partial E}{\partial w_{ij}^{(l)}} = a_j^{(l-1)} \delta_i^{(l)} \tag{15}$$

$$\frac{\partial E}{\partial b_i^{(l)}} = \delta_i^{(l)} \tag{16}$$

where $\delta^{(L)}$ is the error vector at the output layer, $J^{(l)}$ is the Jacobian matrix of layer l, μ is a regularization parameter that controls the step size, and I is the identity matrix. The derivative of the cost function with respect to the weights and biases is calculated as follows:

$$\frac{\partial E}{\partial w_{ij}^{(l)}} = \frac{\partial E}{\partial h_i^{(l)}} \frac{\partial h_i^{(l)}}{\partial w_{ij}^{(l)}} \tag{17}$$

$$= \delta_i^{(l)} a_j^{(l-1)} \tag{18}$$

$$\frac{\partial E}{\partial b_i^{(l)}} = \frac{\partial E}{\partial h_i^{(l)}} \frac{\partial h_i^{(l)}}{\partial b_i^{(l)}} \tag{19}$$

$$= \delta_i^{(l)} \tag{20}$$

The weights and biases are then updated using the following rules in each iteration:

$$w_{ij}^{(l)} := w_{ij}^{(l)} - \eta \frac{\partial E}{\partial w_{ij}^{(l)}} \tag{21}$$

$$b_i^{(l)} := b_i^{(l)} - \eta \frac{\partial E}{\partial b_i^{(l)}} \tag{22}$$

Here, η denotes the learning rate, which controls the step size of the weight and bias updates. The algorithm is repeatedly applied until the error converges or a maximum number of epochs is reached.

This study uses feedforward ANN with Levenberg-Marquardt backpropagation as described in equations (1)–(22). Table 1 shows the values of the ANN parameters used, including the number of inputs, number of outputs, number of hidden layers and neurons, the maximum number of epochs, maximum training time, and performance goal.

2.2.2. Gradient boosting (GB)

Gradient boosting is a powerful machine learning technique that iteratively combines multiple weak models to create a strong model [35]. Fig. 3 shows the GB algorithm utilized in this study. The algorithm can be broken down into three main steps, and three substeps.

The first main step is model initialization, in this step, the model is initialized with a constant random value that minimizes the loss function (L) as in equation (23).

Table 1
ANN hyperparameter values and descriptions.

Parameter	Description	Value
Number of inputs	Number of input data variables	9
Number of outputs	Number of output forecasted variables	1
Number of hidden layers	Number of hidden layers hyperparameter	2
Number of hidden neurons	Number of hidden neurons hyperparameter	9
Maximum epochs	Max. number of training iterations before training is stopped	1000
Maximum training time	Max. time in seconds before training is stopped	Unlimited
Performance goal	The minimum target value of MSE	0

$$F_0(x) = \arg \gamma \sum_{i=1}^{n} L(y_i, \gamma) \tag{23}$$

where the loss function $L(y_i, \gamma)$ measures the difference between y_i (observed value) and γ (predicted value) from the first till the $i-th$ observation.

After the model is initialized, the second main step (model fitting) is performed. In this step, the model is fitted by iterating through M number of trees, starting from the first iteration ($m=1$), till M. This step consists of three substeps:

1. Compute pseudo-residuals (r) for the $m-th$ iteration as in equation (24).

$$\boldsymbol{r_{im}} = -\left[\frac{\partial \boldsymbol{L}(\boldsymbol{y_i}, \boldsymbol{F}(\boldsymbol{x_1}))}{\partial \boldsymbol{F}(\boldsymbol{x_i})}\right]_{\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{F_{m-1}}(\boldsymbol{x}) - \mathbf{1}(\boldsymbol{x})} \tag{24}$$

where x is the input feature vector for a specific data point (i). -$\left[\frac{\partial L(y_i, F(x_1))}{\partial F(x_i)}\right]$ is the negative gradient of the loss function. While $F_{m-1}(x) - 1(x)$ is the previous iteration's prediction.

2. Fit regression tree (T) to the pseudo-residuals by finding the structure that minimizes the loss function as in equation (25).

$$T_m(x) = \operatorname{argmin}(\alpha, j, s) \sum_{i=1}^{n} L\left(y_i, F_{m-1}(x_i) + \alpha I\left(x_{ij} \le s\right)\right) \tag{25}$$

where x_{ij} is the value of the $j-th$ feature for the $i-th$ observation. $I(x_{ij} \le s)$ is the penalty term which penalizes splits that do not improve the prediction accuracy. Note that to minimize the loss function L, the values of α, j, and s must be optimized. As the α value is the prediction assigned to each leaf of the tree, while j and s refer to the feature and feature value, the structure of the regression tree is determined by the values of these three parameters. Using the optimized values of α, j, and s, a regression tree can be built to determine the step size γ_m for the current iteration (m) of the algorithm as in equation (26).

$$\gamma_m = \arg(\gamma) \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \gamma T_m(x_i)) \tag{26}$$

The step size γ_m is determined by minimizing the loss function for the current iteration, which includes the model (F_{m-1}). The new weak learner (T_m) is scaled by γ which is a hyperparameter that is determined by minimizing the loss function.

3. Update the model by adding the prediction of the current tree scaled by the step size as in equation (27):

$$F_m(x) = F_{m-1}(x) + \gamma_m T_m(x) \tag{27}$$

Finally, in the third main step, the prediction for a new observation

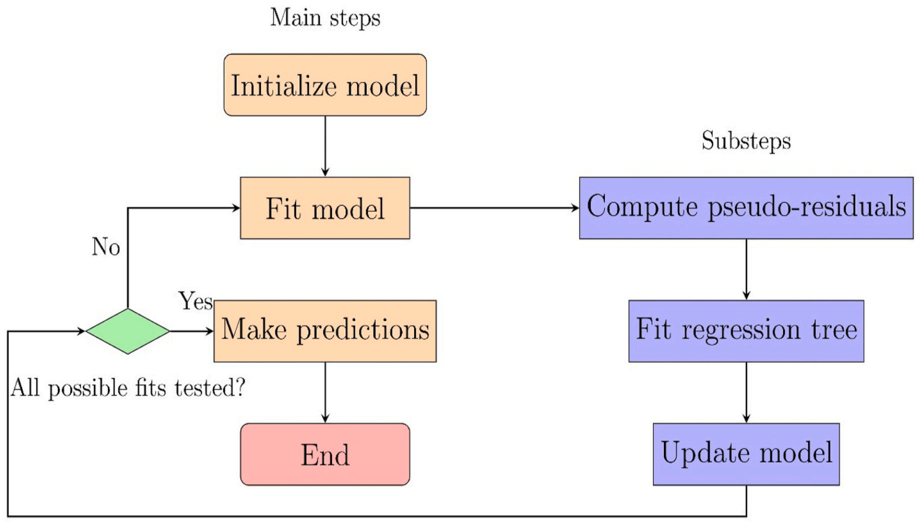

Fig. 3. Gradient boosting algorithm workflow.

$\hat{y}_i$ is performed as in equation (28):

$$\hat{y}_i = F_M(x_i) \tag{28}$$

The final prediction of the gradient boosting regressor model in equation (28) is obtained by using the ensemble of trees that were optimized by minimizing the loss function. The algorithm can improve the model's performance by testing different hyperparameter combinations to find the optimal set. Table 2 describes the parameters used in optimizing the GB models. The Huber loss function is chosen to be optimized. The learning rate, which controls the step size at each iteration, is tested at 8 different values ranging from 0.05 to 0.40, while the number of boosting stages (number of estimators) and the maximum depth of the individual regression estimators (max depth) are both tested at 8 different values ranging from 25 to 200 and 2 to 9, respectively. To evaluate the performance of each combination of hyperparameters, the model is trained using a K-fold cross-validation technique, with 3 folds used for each of the 512 candidates (3 parameters each with 8 different values 8^3, as shown in Table 2). In total, 1536 fits (512*3) are carried out for each model (three different models were built for the three regions, USA, China, and EU).

3. Results

In the following sub-sections, the results start with the comparative descriptives for all investigated regions, followed by the machine learning results laid separately for each geography (i.e. USA, China, and EU).

3.1. Regional comparisons

Regarding the historical GHG emissions data, Fig. 4 shows GHG emissions in the USA, China, and the EU from 1965 until 2021. It can be observed that China experienced a significant increase in emissions over the last two decades. Conversely, the USA and the EU have shown a slight decrease in emissions.

Fig. 5 shows the proportions of energy consumption for each type of energy source in the USA, China, and the EU for the year 2021. It can be noted from the graph that the three major economies extremely depend on fossil fuels. USA and EU exhibit a similar consumption pattern as both rely mostly on oil, followed by gas then coal to supply their demand. However, the EU has a relatively higher utilization of renewable resources compared to the USA. China on the other hand shows a substantial dependence on coal, with over 55% of energy consumption coming from coal. The current situation (i.e. 2021 numbers) is used as a reference point for the effects of increasing the share of energy consumption sources on GHG emissions.

Table 2
Gradient Boosting regressor hyperparameter values and descriptions.

Parameter	Description	Tested values during optimization
Loss	The loss function to be optimized	Huber
Learning rate	The learning rate that controls the step size at each iteration while moving toward the minimum loss function	0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.40
Number of estimators	The number of boosting stages to be performed	25, 50, 75, 100, 125, 150, 175, 200
Max depth	Maximum depth of the individual regression estimators	2, 3, 4, 5, 6, 7, 8, 9

3.2. USA results

3.2.1. Gradient boosting results for the USA

Based on the optimization technique, the best parameters found for the US GB model were a learning rate of 0.05, a maximum depth of 2, and 200 estimators. These parameters resulted in a 0.95 r-squared value, indicating that the model is a good fit for the data.

The GB results for the US are depicted in Fig. 6. Energy consumption from coal plays the highest importance (0.41) in GHG emissions, followed by nuclear (0.23), oil (0.22), wind (0.07), geo biomass (0.04), and solar (0.03). Gas, hydro, and biofuels have very marginal effects. The results for the USA demonstrate that the GHG emissions can be mainly explained by coal, nuclear, and oil consumption. Coal, nuclear, and oil consumption sources were further analyzed using ANN to investigate the effects of increasing the consumption shares of each of these sources on GHG emissions.

3.2.2. ANN results for the USA

The USA ANN model had an r-squared value of 0.98, indicating that the model is a good fit for the data. Fig. 7 shows the effects of increasing coal consumption on GHG emissions in the USA where it shows a positive relationship between coal consumption and GHG emissions. Increasing coal consumption by 25% will result in a 13% increase in GHG emissions.

Fig. 8 shows the effects of increasing nuclear consumption on GHG emissions in the USA. The model shows a positive relationship between nuclear consumption and GHG emissions. Increasing nuclear consumption by 25% will result in about 6% increase in GHG emissions.

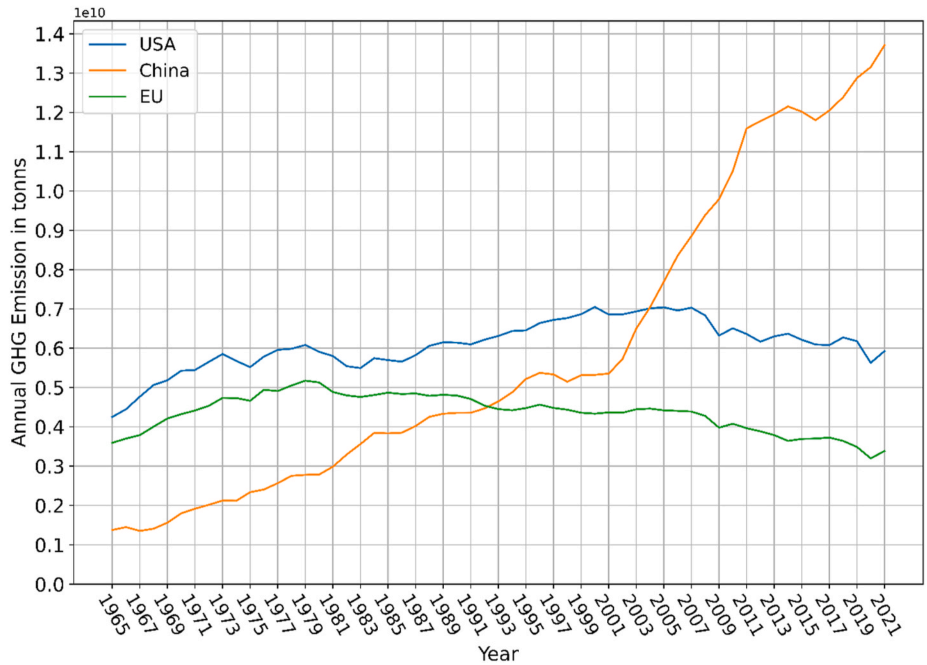

Fig. 4. Comparison of historical GHG Emission in USA, China, and EU

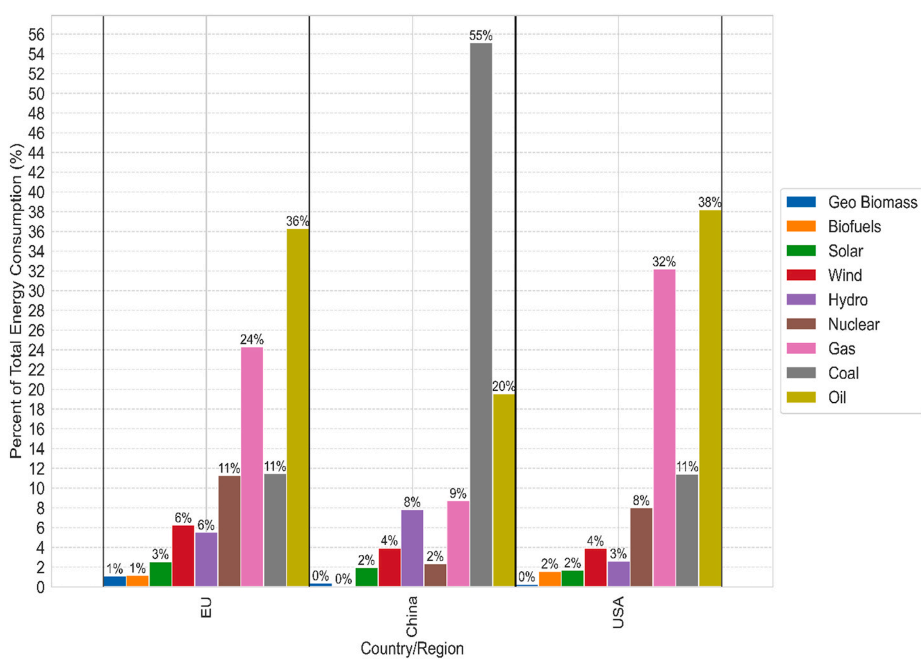

Fig. 5. Comparison of energy consumption by type for USA, China, and EU in 2021.

Similarly, Fig. 9 shows the effects of increasing the US oil consumption on GHG emissions. Increasing oil consumption by 25% will result in almost an 8.4% increase in GHG emissions.

3.3. China results

3.3.1. Gradient boosting results for China

Based on the optimization technique, the best parameters found for the Chinese GB model were a learning rate of 0.05, a maximum depth of 5, and 200 estimators. These parameters resulted in 0.98 r-squared value, indicating that the model is a good fit for the data.

The GB results for China are depicted in Fig. 10. Energy consumption from nuclear plays the highest importance (0.21) in GHG emissions, followed by wind (0.17), gas (0.15), geo biomass (0.13), oil (0.12), hydro (0.09), coal (0.07), solar (0.05), and lastly biofuels (0.02). The results for China demonstrate that the GHG emissions can be mainly explained by nuclear, wind, and gas consumption. Those will be further analyzed using ANN to investigate the effects of increasing the consumption shares of each of these sources and GHG emissions.

3.3.2. ANN results for China

The Chinese ANN model had an r-squared value of 0.99, indicating that the model is a good fit for the data. Fig. 11 shows the effects of increasing nuclear consumption on GHG emissions in China. The model

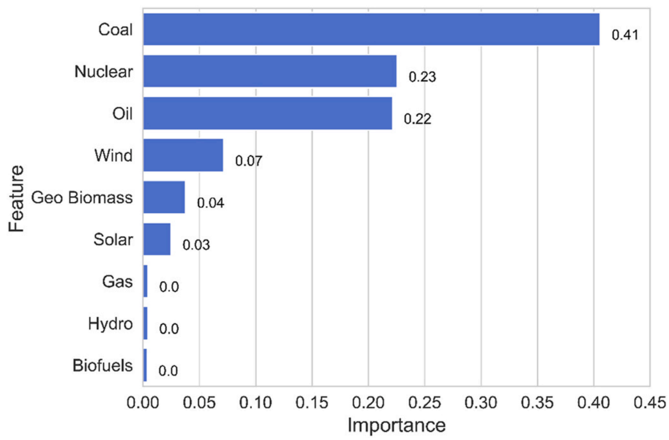

Fig. 6. GB feature importance for the USA.

shows a negative relationship between nuclear consumption and GHG emissions. Increasing nuclear consumption by 25% will result in an 11% decrease in GHG emissions. Fig. 12 shows the effects of increasing wind consumption on GHG emissions in China as predicted by the ANN model. The model shows a negative relationship between wind consumption and GHG emissions. Increasing wind consumption by 25% will result in about 3% decrease in GHG emissions. Similarly, Fig. 13 shows the effects of increasing gas consumption on GHG emissions in China. The model shows a positive relationship between gas consumption and GHG emissions and increasing gas consumption by 25% will result in about 11% increase in GHG emissions.

3.4. EU results

3.4.1. Gradient boosting results for the EU

Based on the optimization technique, the best parameters found for the EU GB model were a learning rate of 0.01, a maximum depth of 2, and 100 estimators. These parameters resulted in 0.97 r-squared value, indicating that the model is a good fit for the data.

The GB model analysis for the EU shows that energy consumption from oil plays the highest importance (0.33) in GHG emissions, followed by coal (0.22), wind (0.13), biofuels (0.13), nuclear (0.10), geo biomass (0.05) and gas (0.04). While solar and hydro have negligible effects. The results for the EU demonstrate that the GHG emissions can be mainly explained by oil, coal, and wind energy consumption as shown in Fig. 14. Those will be further analyzed using ANN to investigate the effects of increasing the consumption shares of each of these sources and GHG emissions.

3.4.2. ANN results for the EU

Fig. 15 shows the effects of increasing oil consumption on GHG emissions in the EU. The model shows that increasing oil consumption has a minor negative effect on GHG emissions. Increasing oil consumption by 25% will result in a 1% decrease in GHG emissions.

Fig. 16 shows the effect of increasing coal consumption on GHG emissions in the EU; the model shows a positive relationship between coal consumption and GHG emissions where increasing coal consumption by 25% will result in about an 11% increase in GHG emissions. Similarly, Fig. 17 shows the effects of increasing wind consumption on GHG emissions in the EU; the model shows increasing wind power consumption has a minor effect on GHG emissions. Where increasing wind consumption by 25% will result in about a 2% increase in GHG emissions.

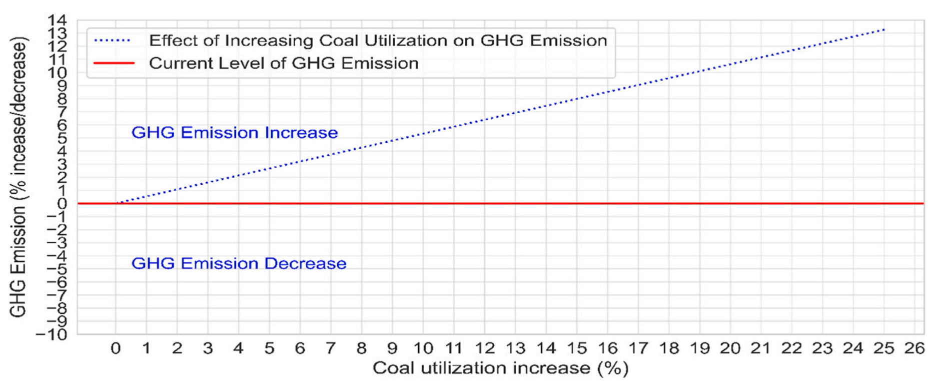

Fig. 7. Effect of increasing coal consumption on GHG in the USA.

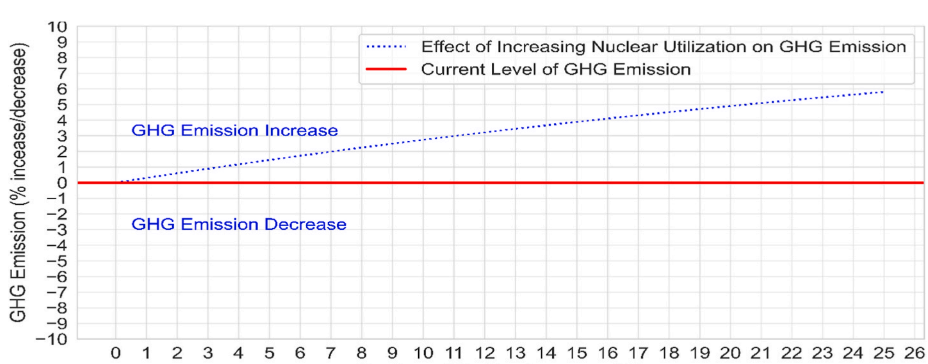

Fig. 8. Effect of increasing nuclear consumption on GHG in the USA.

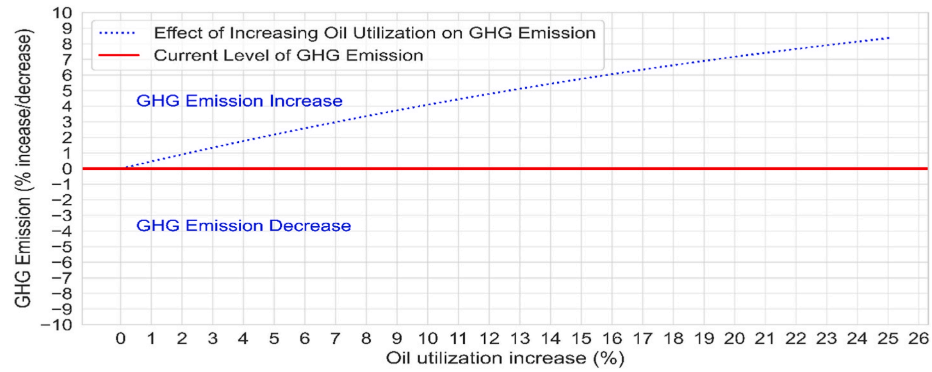

Fig. 9. Effect of increasing oil consumption on GHG in the USA.

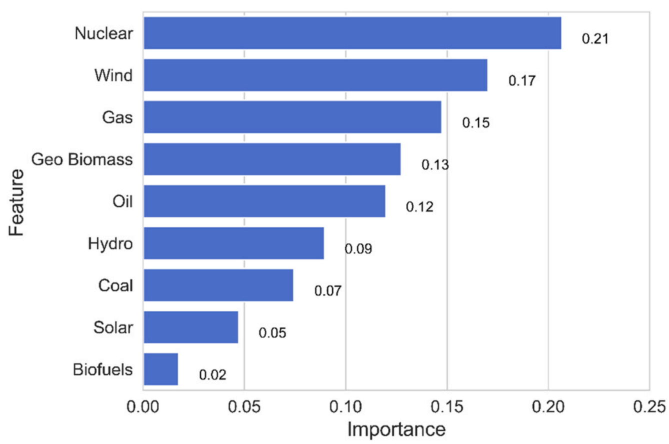

Fig. 10. GB feature importance for China.

4. Discussion

This study aimed to investigate the major sources contributing to greenhouse gas (GHG) emissions in the USA, China, and the EU using an ensemble tree machine learning model known as GB and ANN. The GB model was preferred for its ability to identify weaker non-linear relationships that could better explain the target variable. The GB results revealed that the major sources contributing to GHG emissions in the USA are energy consumption from coal, nuclear, and oil. For China, GB demonstrated that nuclear, wind, and gas consumption were the most significant factors contributing to GHG emissions. While for the EU, the major factors contributing to GHG emissions were found to be energy consumption from oil, coal, and wind sources. We chose to use an ensemble tree machine learning model such as GB, as it is possible for some sources that have a high usage rate, and still not to be identified as one of the most important in the feature analysis. This is due to the functionality of the GB models which is based on how much each feature contributes to reducing errors in predicting the target variable, rather than its current weight [36]. This suggests that while some features may have strong linear relationships with the target variable, other features with weaker non-linear relationships may better explain the target variable [37–39]. This is particularly important in China where coal consumption accounted for 55% of total energy consumption, yet it is not among the most important factor affecting GHG emissions. Nevertheless, smaller utilized sources such as nuclear (2% of Chinese energy consumption) are among the most important sources impacting GHG emissions. Similarly, wind energy utilization is smaller than other sources in USA and EU, however, it is among the most important feature of explaining GHG emissions.

In parallel, the results of the ANN models for the USA showed that increasing coal consumption by 25% would result in a 13% increase in GHG emissions. This suggests that reducing coal consumption could be the most effective strategy to reduce GHG emissions in the USA. Meanwhile, increasing nuclear consumption by 25% would result in about a 6% increase in GHG emissions, which is lower than the increase seen with coal consumption. However, this also suggests that increasing nuclear consumption may not be an effective strategy for reducing GHG emissions, contrary to what other researchers recommend [40,41]. Moreover, it was found that increasing oil consumption by 25% would result in about an 8.4% increase in GHG emissions which suggests that reducing oil consumption could also be an effective strategy for reducing GHG emissions in the USA. However, the priority for the USA should be targeted towards reducing coal consumption as it has the highest effect on GHG emissions.

Regarding the Chinese ANN model results, it was found that increasing nuclear consumption by 25% will result in a significant 11% decrease in GHG emissions. Therefore, increasing nuclear energy

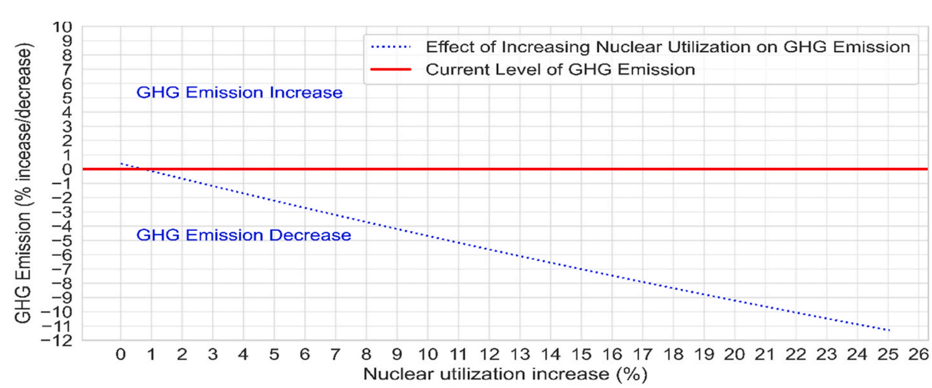

Fig. 11. Effect of increasing nuclear consumption on GHG in China.

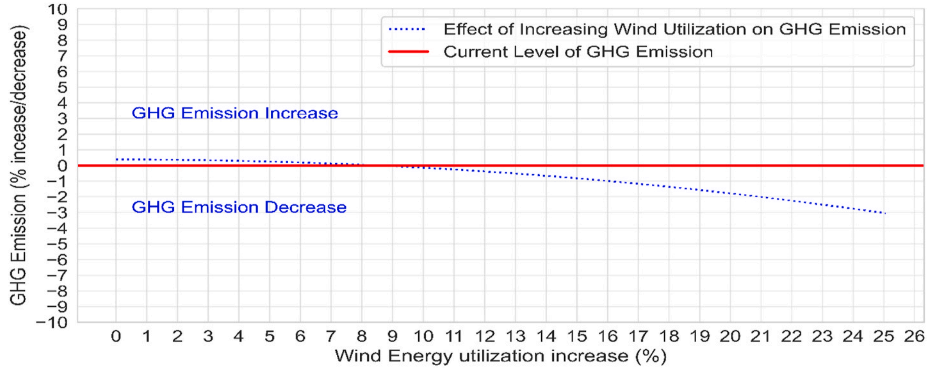

Fig. 12. Effect of increasing wind consumption on GHG in China.

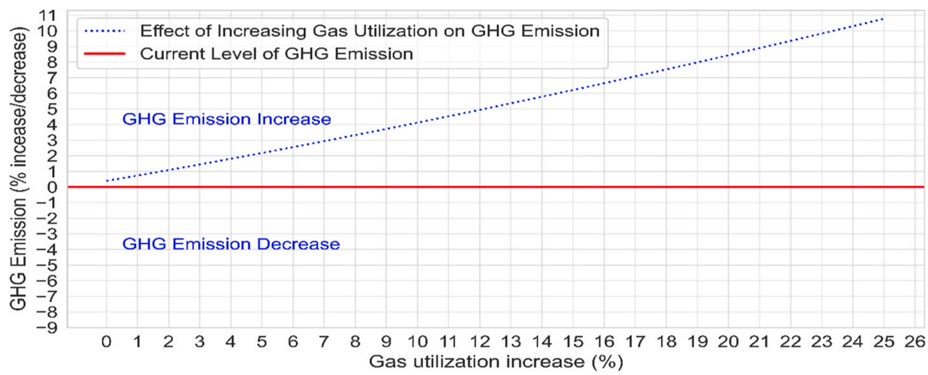

Fig. 13. Effect of increasing gas consumption on GHG in China.

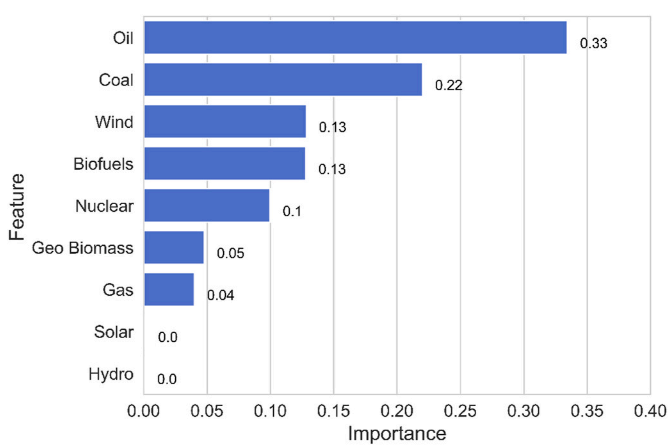

Fig. 14. GB feature importance for EU

production can contribute to reducing GHG emissions in China, contrary to the USA. Similarly, increasing wind consumption by 25% will result in about a 3% decrease in GHG emissions in China. which suggests that wind energy is also a significant low-carbon energy source that can contribute to reducing GHG emissions. In contrast, increasing gas consumption by 25% will result in about an 11% increase in GHG emissions, which suggests that reducing gas consumption could also be an effective strategy for reducing GHG emissions in China.

The paradox of increasing nuclear energy consumption being coupled with directly decreasing GHG emissions in China can be attributed to the displacement of fossil fuel-based energy sources, particularly coal, as China is heavily reliant on coal as the primary source of energy (55% of total energy sources). Additionally, China's nuclear power plants are relatively new and efficient, with fewer backups from fossil fuel sources compared to the USA [42,43]. This displacement effect has been observed in previous studies [44]. On the other, the positive relationship between nuclear consumption and GHG emissions in the USA can be attributed to the fact that the nuclear power plants in the USA are relatively old and are often backed up by coal-fired power plants, which in turn emit a significant amount of GHG emissions [45].

The results of the ANN models for the EU concluded that increasing oil consumption has a minor negative effect on GHG emissions. This emphasizes that the EU has successfully turned oil into a carbon-neutral source with a very marginal carbon footprint compared to other sources. The EU policies and regulations to reduce GHG emissions from oil sources, especially in the transport sector as well as improving the oil-fired power plants' efficiency were the main drivers for this success [46,47]. On the other hand, the model shows a positive relationship between coal consumption and GHG emissions in the EU, indicating that increasing coal consumption by 25% will result in an 11% increase in GHG emissions. This finding highlights the importance of reducing coal consumption to decrease GHG emissions in the EU. It was also found that increasing wind energy consumption by 25% will surprisingly result in an increase of about 2% in GHG emissions; this can be attributed to two reasons, one is that the technology used is still inefficient, and the other is that the location selection should be optimized. This suggests that the EU is unlikely to environmentally benefit from expanding its wind energy base using the current technologies and settings. The conclusion that wind energy has a mixed impact on GHG emissions as observed in the EU and China is corroborated by several empirical studies. While some research has found no evidence to support the claim that wind energy can mitigate global warming, other studies have observed a positive impact on environmental quality through the reduction of GHG emissions [48].

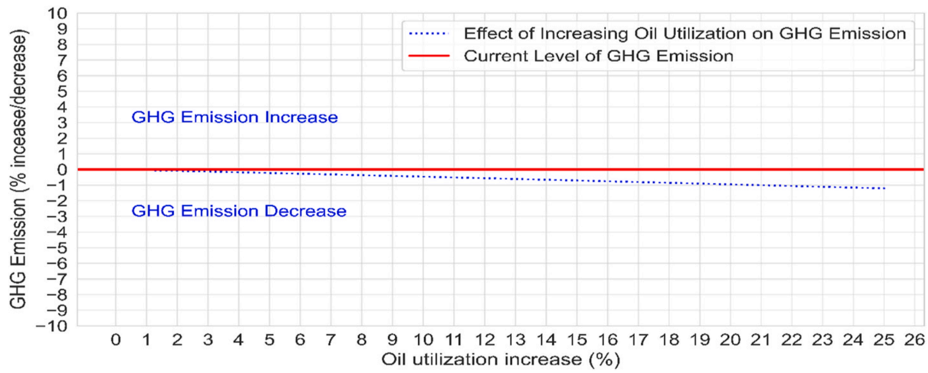

Fig. 15. Effect of increasing oil consumption on GHG in the EU

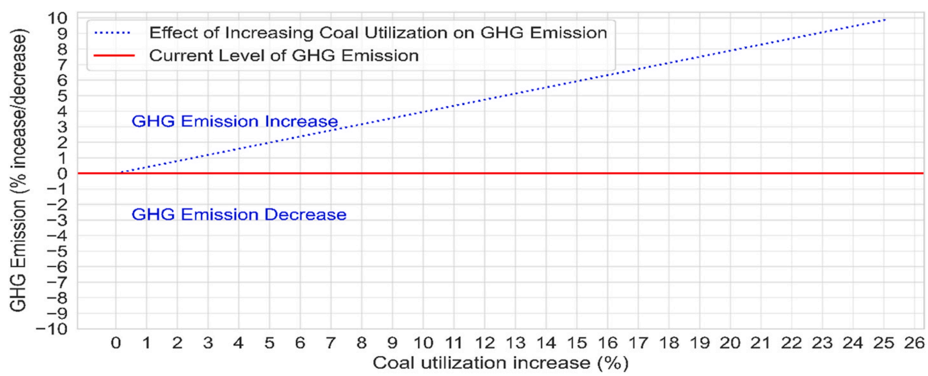

Fig. 16. Effect of increasing coal consumption on GHG in the EU

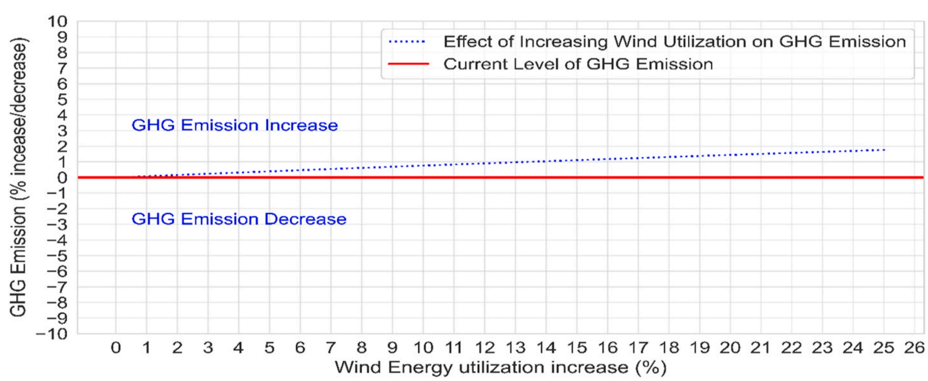

Fig. 17. Effect of increasing wind consumption on GHG in the EU

5. Conclusion and policy recommendations

This study utilizes machine learning techniques such as ANN and GB to provide novel insights into the energy consumption sources that drive GHG in the USA, China, and the EU, three major economies and global players in terms of energy consumption and emissions. Energy consumption and GHG emissions data were obtained from “Our World in Data” database covering the period from 1965 to 2021. The findings have significant implications for policymakers and suggest that effective GHG emissions reduction strategies must be tailored to the specific energy utilization sources of each country. For the USA, the priority should be on reducing coal consumption. It was also found that increasing nuclear consumption may not be an effective strategy for the US. In China, increasing nuclear energy production and wind consumption can contribute to reducing GHG emissions. While in the EU, policies and regulations should be implemented to reduce GHG emissions from coal sources, as well as improve efficiency and optimize location selection for wind turbines.

Future work can focus on addressing a number of limitations in this study. The accuracy and reliability of the analysis depend on the availability and quality of data. Especially since our source utilizes the IPCC methodology in data collection. Although the IPCC methodology is based on a rigorous review of the scientific literature and undergoes an extensive review process to ensure its accuracy and completeness, however, like any scientific endeavor, it is subject to limitations. One of the limitations is that this method is based on conversion factors which might create some degree of uncertainty in the data. Also, the complexity and Interpretability of the ANN and gradient boosting models are multifaceted. While these black-box models can capture intricate relationships between input variables, interpreting the results and understanding the underlying mechanisms can be challenging.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT and Bing in order to improve readability and language. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] C.B. Field, V.R. Barros, Climate Change 2014–Impacts, Adaptation and Vulnerability: Regional Aspects, Cambridge University Press, 2014.
[2] J.A. Patz, H.K. Gibbs, J.A. Foley, J.V. Rogers, K.R. Smith, Climate change and global health: quantifying a growing ethical crisis, EcoHealth 4 (2007) 397–405.
[3] C. Albergel, J.-C. Calvet, A.-L. Gibelin, S. Lafont, J.-L. Roujean, C. Berne, O. Traullé, N. Fritz, Observed and modelled ecosystem respiration and gross primary production of a grassland in southwestern France, Biogeosciences 7 (5) (2010) 1657–1668.
[4] S. Kuşkaya, F. Bilgili, E. Muğaloğlu, K. Khan, M.E. Hoque, N. Toguç, The role of solar energy usage in environmental sustainability: fresh evidence through time-frequency analyses, Renew. Energy 206 (2023) 858–871.
[5] A. Mikhaylov, N. Moiseev, K. Aleshin, T. Burkhardt, Global climate change and greenhouse effect, Entrepreneurship and Sustainability Issues 7 (4) (2020) 2897.
[6] S. Kuşkaya, Residential solar energy consumption and greenhouse gas nexus: evidence from Morlet wavelet transforms, Renew. Energy 192 (2022) 793–804.
[7] G. Janssens-Maenhout, M. Crippa, D. Guizzardi, M. Muntean, E. Schaaf, F. Dentener, P. Bergamaschi, V. Pagliari, J.G. Olivier, J.A. Peters, EDGAR v4. 3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012, Earth Syst. Sci. Data 11 (3) (2019) 959–1002.
[8] N. Xu, S. Ding, Y. Gong, J. Bai, Forecasting Chinese greenhouse gas emissions from energy consumption using a novel grey rolling model, Energy 175 (2019) 218–227.
[9] F.W. Geels, F. Berkhout, D.P. Van Vuuren, Bridging analytical approaches for low-carbon transitions, Nat. Clim. Change 6 (6) (2016) 576–583.
[10] H. Haberl, D. Wiedenhofer, D. Virág, G. Kalt, B. Plank, P. Brockway, T. Fishman, D. Hausknost, F. Krausmann, B. Leon-Gruchalski, A systematic review of the evidence on decoupling of GDP, resource use and GHG emissions, part II: synthesizing the insights, Environ. Res. Lett. 15 (6) (2020), 065003.
[11] D.I. Stern, Energy-GDP relationship, The New Palgrave dictionary of economics (2018) 3697–3714.
[12] Y.-J. Zhang, Y.-B. Da, The decomposition of energy-related carbon emission and its decoupling with economic growth in China, Renew. Sustain. Energy Rev. 41 (2015) 1255–1266.
[13] A. Olabi, M.A. Abdelkareem, Renewable energy and climate change, Renew. Sustain. Energy Rev. 158 (2022), 112111.
[14] M. Croucher, Potential problems and limitations of energy conservation and energy efficiency, Energy Pol. 39 (10) (2011) 5795–5799.
[15] E. Eaton, Approaches to energy transitions: carbon pricing, managed decline, and/or green new deal? Geography Compass 15 (2) (2021), e12554.
[16] O. Edenhofer, Climate Change 2014: Mitigation of Climate Change, Cambridge University Press, 2015.
[17] K. Sadamori, Nuclear power in a clean energy system, in: Annales des Mines-Responsabilité et environnement, Cairn/Softwin, 2020, pp. 122–126.
[18] M. Lenzen, Life cycle energy and greenhouse gas emissions of nuclear energy: a review, Energy Convers. Manag. 49 (8) (2008) 2178–2199.
[19] P.R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.O. Pörtner, D. Roberts, P. Zhai, R. Slade, S. Connors, R. Van Diemen, IPCC, 2019: Climate Change and Land: an IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems, 2019.
[20] K. Liu, Z. Wei, Z. Yang, K. Li, Mass load prediction for lithium-ion battery electrode clean production: a machine learning approach, J. Clean. Prod. 289 (2021), 125159.
[21] R. Olu-Ajayi, H. Alaka, Building Energy Consumption Prediction Using Deep Learning, Environmental Design and Management Conference (EDMIC), 2021.
[22] D.A. Otchere, T.O.A. Ganat, R. Gholami, S. Ridha, Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: comparative analysis of ANN and SVM models, J. Petrol. Sci. Eng. 200 (2021), 108182.
[23] O.I. Abiodun, A. Jantan, A.E. Omolara, K.V. Dada, N.A. Mohamed, H. Arshad, State-of-the-art in artificial neural network applications: a survey, Heliyon 4 (11) (2018), e00938.
[24] L.C. Assessment, Life cycle greenhouse gas emissions from electricity generation: update, Life 800 (1) (2021).
[25] O. Edenhofer, R. Pichs-Madruga, Y. Sokona, K. Seyboth, P. Matschoss, S. Kadner, T. Zwickel, P. Eickemeier, G. Hansen, S. Schlömer, IPCC special report on renewable energy sources and climate change mitigation, in: Prepared by Working Group III of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK, 2011.
[26] M.Z. Jacobson, Review of solutions to global warming, air pollution, and energy security, Energy Environ. Sci. 2 (2) (2009) 148–173.
[27] S. Sorrell, J. Dimitropoulos, M. Sommerville, Empirical estimates of the direct rebound effect: a review, Energy Pol. 37 (4) (2009) 1356–1371.
[28] S.M. Benson, F.M. Orr, Carbon dioxide capture and storage, MRS Bull. 33 (4) (2008) 303–305.
[29] B. Nykvist, M. Nilsson, Rapidly falling costs of battery packs for electric vehicles, Nat. Clim. Change 5 (4) (2015) 329–332.
[30] R. Deng, Z. Yang, M.-Y. Chow, J. Chen, A survey on demand response in smart grids: mathematical models and approaches, IEEE Trans. Ind. Inf. 11 (3) (2015) 570–582.
[31] A. Shojaeighadikolaei, A. Ghasemi, K. Jones, Y. Dafalla, A.G. Bardas, R. Ahmadi, M. Haashemi, Distributed Energy Management and Demand Response in Smart Grids: A Multi-Agent Deep Reinforcement Learning Framework, 2022 arXiv preprint arXiv:2211.15858.
[32] J. Yu, K. Spiliopoulos, Normalization Effects on Deep Neural Networks, 2022 arXiv preprint arXiv:2209.01018.
[33] M.W. Jones, G.P. Peters, T. Gasser, R.M. Andrew, C. Schwingshackl, J. Gütschow, R.A. Houghton, P. Friedlingstein, J. Pongratz, C. Le Quéré, National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850, Sci. Data 10 (1) (2023) 155.
[34] Q. Ahmad, R.A. Warrick, T. Downing, S. Nishioka, K. Parikh, C. Parmesan, S. Schneider, F. Toth, G. Yohe, Methods and Tools. Climate Change, 2001, pp. 105–143.
[35] Friedman, J.H.J.A.o.s., Greedy Function Approximation: a Gradient Boosting Machine, 2001, pp. 1189–1232.
[36] J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2009 (No Title).
[37] G. Casalicchio, C. Molnar, B. Bischl, Visualizing the Feature Importance for Black Box Models, Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I 18, Springer, 2019, pp. 655–670.
[38] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[39] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[40] S. Lee, M. Kim, J. Lee, Analyzing the impact of nuclear power on CO2 emissions, Sustainability 9 (8) (2017) 1428.
[41] T. Li, D. Du, X. Wang, X. Qin, Can nuclear power products mitigate greenhouse gas emissions? Evidence from global trade network, Int. J. Environ. Res. Publ. Health 19 (13) (2022) 7808.
[42] B.W. Brook, A. Alonso, D.A. Meneley, J. Misak, T. Blees, J.B. van Erp, Why nuclear energy is sustainable and has to be part of the energy mix, Sustainable Materials and Technologies 1 (2014) 8–16.
[43] B. Lin, Y. Xie, Analysis on operational efficiency and its influencing factors of China's nuclear power plants, Energy 261 (2022), 125211.
[44] R. Ulucak, S. Erdogan, The effect of nuclear energy on the environment in the context of globalization: consumption vs production-based CO2 emissions, Nucl. Eng. Technol. 54 (4) (2022) 1312–1320.
[45] C. Magazzino, M. Mele, C. Drago, S. Kuşkaya, C. Pozzi, U. Monarca, The trilemma among CO2 emissions, energy use, and economic growth in Russia, Sci. Rep. 13 (1) (2023), 10225.
[46] M. Galeotti, M. Manera, A. Lanza, On the robustness of robustness checks of the environmental Kuznets curve hypothesis, Environ. Resour. Econ. 42 (2009) 551–574.
[47] U. Al-Mulali, C.F. Tang, I. Ozturk, Estimating the environment Kuznets curve hypothesis: evidence from Latin America and the Caribbean countries, Renew. Sustain. Energy Rev. 50 (2015) 918–924.
[48] S. Kuşkaya, F. Bilgili, The wind energy-greenhouse gas nexus: the wavelet-partial wavelet coherence model approach, J. Clean. Prod. 245 (2020), 118872.