

Article

A Comprehensive Approach to CO₂ Emissions Analysis in High-Human-Development-Index Countries Using Statistical and Time Series Approaches

Hamed Khosravi, Ahmed Shoyeb Raihan , Farzana Islam, Ashish Nimbarte  and Imtiaz Ahmed * 

Department of Industrial and Management Systems Engineering, West Virginia University, Morgantown, WV 26506, USA; hamed.khosravi@mail.wvu.edu (H.K.); ar00065@mix.wvu.edu (A.S.R.); fi00003@mix.wvu.edu (F.I.); ashish.nimbarte@mail.wvu.edu (A.N.)

* Correspondence: imtiaz.ahmed@mail.wvu.edu

Abstract: Reducing carbon dioxide (CO₂) emissions is vital at both global and national levels, given their significant role in exacerbating climate change. CO₂ emissions, stemming from a variety of industrial and economic activities, are major contributors to the greenhouse effect and global warming, posing substantial obstacles in addressing climate issues. It is imperative to forecast CO₂ emissions trends and classify countries based on their emission patterns to effectively mitigate worldwide carbon emissions. This paper presents an in-depth comparative study on the determinants of CO₂ emissions in twenty countries with high Human Development Index (HDI), exploring factors related to economy, environment, energy use, and renewable resources over a span of 25 years. The study unfolds in two distinct phases: initially, statistical techniques such as Ordinary Least Squares (OLS), fixed effects, and random effects models are applied to pinpoint significant determinants of CO₂ emissions. Following this, the study leverages supervised and unsupervised time series approaches to further scrutinize and understand the factors influencing CO₂ emissions. Seasonal AutoRegressive Integrated Moving Average with exogenous variables (SARIMAX), a statistical time series forecasting model, is first used to predict emission trends from historical data, offering practical insights for policy formulation. Subsequently, Dynamic Time Warping (DTW), an unsupervised time series clustering approach, is used to group countries by similar emission patterns. The dual-phase approach utilized in this study significantly improves the accuracy of CO₂ emissions predictions while also providing a deeper insight into global emission trends. By adopting this thorough analytical framework, nations can develop more focused and effective carbon reduction policies, playing a vital role in the global initiative to combat climate change.

Keywords: CO₂ emissions; carbon footprint reduction; emission trend forecasting; CO₂ indicators; time series approaches; sustainable future



Academic Editor: Antonio Caggiano

Received: 12 December 2024

Revised: 3 January 2025

Accepted: 5 January 2025

Published: 14 January 2025

Citation: Khosravi, H.; Raihan, A.S.; Islam, F.; Nimbarte, A.; Ahmed, I. A Comprehensive Approach to CO₂ Emissions Analysis in High-Human-Development-Index Countries Using Statistical and Time Series Approaches. *Sustainability* **2025**, *17*, 603. <https://doi.org/10.3390/su17020603>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Climate change stands as one of the most formidable challenges of our time, widely recognized as an unequivocal and potentially irreversible threat to the sustainability of our global environment [1]. The consensus among scientists and researchers is overwhelming, as highlighted in the Intergovernmental Panel on Climate Change's (IPCC) Sixth Assessment Report (AR6, 2021), which explicitly outlines the critical impacts and long-term implications of climate change on ecosystems, biodiversity, and human societies [2,3]. At the heart of this environmental conundrum lies the pivotal role of CO₂ emissions,

identified as a primary contributor to the greenhouse effect and a key driver of global warming [4]. While this study focuses on CO₂, it is essential to contextualize it within the broader spectrum of greenhouse gases, which include methane (CH₄), nitrous oxide (N₂O), and fluorinated gases. Together, these gases contribute to global warming, with CO₂ accounting for approximately 76% of total GHG emissions, followed by methane at 16% and nitrous oxide at 6% [5,6]. The dominance of CO₂ in the GHG mix underscores its critical role in global emissions reduction efforts, making it the focal point of this study. CO₂ emissions result predominantly from the burning of fossil fuels, such as coal, oil, and natural gas, a practice intricately linked to industrial activities, energy production, and transportation [7,8]. The accumulation of CO₂ in the Earth's atmosphere traps heat, leading to a rise in global temperatures, a phenomenon commonly referred to as global warming. The effects of this warming are far-reaching, ranging from more frequent and severe weather events, such as hurricanes and droughts, to long-term shifts in climate patterns, impacting agriculture, ecosystems, and sea levels worldwide [9]. Recent research has further solidified the understanding of CO₂'s impact on climate change.

Globally, countries are implementing diverse strategies and policies to combat the seriousness of climate change, such as transitioning to renewable energy sources, enhancing energy efficiency, and engaging in international agreements like the Paris Agreement [10,11]. Germany's Energiewende (energy transition) program has significantly increased the share of renewables in its electricity mix, from 6% in 2000 to over 42% in 2020, contributing to a reduction of approximately 300 million metric tons of CO₂ emissions annually [12]. Similarly, Sweden's carbon tax, introduced in 1991 and one of the highest globally, has resulted in an 18% reduction in fossil fuel consumption and a 12% increase in biofuel production, demonstrating the effectiveness of economic incentives in driving the transition to sustainable energy sources [13]. Efforts like the REDD+ (Reducing Emission from Deforestation and Forest Degradation in Developing Countries) programs focus on reducing emissions from deforestation and forest degradation, which account for approximately 11% of total global greenhouse gas emissions, according to studies on land-use change and forestry [14,15]. These approaches, ranging from carbon pricing mechanisms to investments in green technologies, underscore the importance of understanding the various factors influencing CO₂ emissions. This understanding is crucial for tailoring specific, effective strategies, anticipating and mitigating unintended policy consequences, facilitating informed international cooperation, and ensuring the adaptability and longevity of climate change mitigation efforts. In essence, the success of global initiatives to reduce carbon emissions hinges on a nuanced comprehension of the determinants of CO₂ emissions, a cornerstone in designing and implementing impactful environmental policies [16–18]. In the design and implementation of such programs, a critical element is the identification of factors affecting CO₂ emissions. This requires a multifaceted approach, considering not only direct factors like deforestation and fossil fuel use but also indirect drivers such as economic policies, energy consumption patterns, technological advancement, and societal behaviors [19]. A detailed understanding of these factors is essential for developing effective strategies that can significantly reduce emissions [20].

This research paper seeks to unravel the complexities surrounding CO₂ emissions, with a specific focus on countries with high Human Development Index (HDI) scores [21,22]. These countries, often seen as the vanguards of technological advancement and economic prosperity, find themselves at a critical juncture, where their developmental strides are paradoxically intertwined with substantial environmental footprints [23]. The significance of understanding CO₂ emissions in high-HDI countries cannot be overstated. These nations, which have historically powered their growth through extensive use of fossil fuels, are now at the forefront of a global crisis they significantly contributed to [24]. The environmental

policies and technological innovations that these countries adopt or ignore will have far-reaching consequences, not only within their borders but across the globe. Thus, dissecting the fabric of CO₂ emissions in these contexts becomes not only an environmental imperative but also a moral one. High-HDI countries, such as the United States, Canada, and Germany, collectively account for nearly 44% of global CO₂ emissions despite representing only about 16% of the world's population [25,26]. This disparity is driven by energy-intensive lifestyles, industrialization, and resource consumption that characterize high-development regions. In contrast, lower-HDI countries, while comprising a majority of the global population, contribute less than 14% of global emissions [27]. For instance, countries like Bangladesh and Kenya emit less than 1 metric ton of CO₂ per capita annually, compared to over 15 metric tons in high-HDI countries such as the United States and Canada [28]. While emerging lower-HDI economies are important for analyzing global trends, this study aims to address the pressing need for tailored strategies in high-HDI countries, given their disproportionate impact and leadership role in climate action initiatives. The rationale for focusing on high-HDI countries in this study is three-fold. Firstly, these countries have been the primary architects of the current state of global CO₂ emissions [29]. Their industrial revolutions, fueled by carbon-intensive resources, have left an indelible mark on the planet's atmosphere. Understanding their emission patterns is crucial to untangling the historical context of the current climate crisis and in charting a course for global emission reduction strategies. Secondly, high-HDI countries are often at the cutting edge of technological innovation and economic transformation [30]. This positions them uniquely to lead the way in developing and implementing solutions to reduce CO₂ emissions. From renewable energy technologies to policy innovations for carbon reduction, these countries have the potential to set precedents and create models that can be emulated worldwide. Thirdly, the economic structures and lifestyles in high-HDI countries significantly influence global consumption patterns and, by extension, CO₂ emissions [31]. These nations, therefore, have a disproportionate impact on setting global trends. Their commitment to sustainable practices, or lack thereof, can either exacerbate the climate crisis or steer the world towards a more sustainable trajectory. This research aims to contribute to this vital global endeavor by providing a nuanced understanding of the determinants of CO₂ emissions in high-HDI countries, thereby laying the groundwork for informed, effective, and sustainable policy interventions.

The intricate relationship between CO₂ emissions and various socio-economic, environmental, and energy-related factors has been a focal point of numerous studies over the past decades [32–35]. These research endeavors have collectively contributed to a growing understanding of the multifaceted nature of CO₂ emissions. From an economic perspective, a substantial body of research has explored the relationship between emission and variables like GDP growth, levels of industrialization, and urbanization trends [36–38]. Environmental studies have explored the impact of land use changes, deforestation, and climate policies [39–41]. In terms of energy factors, the role of energy consumption, the mix of energy sources, and the efficiency of energy use have been thoroughly scrutinized [42–44]. Despite the breadth of the existing research, there still remain notable gaps and limitations. One main limitation is the predominant focus on cross-sectional data analysis, which often fails to capture the dynamic nature of the relationship between CO₂ emissions and their determinants over time [45]. Additionally, many studies have been constrained by the scope of their geographical focus, often limiting their analysis to a single country or a small group of countries, thereby restricting the generalizability of their findings [46–49]. Another notable gap is the insufficient exploration of the interaction effects between different determinants of CO₂ emissions, a complexity that demands a more nuanced analytical approach [50–52].

By employing a combination of advanced econometric models and innovative analysis techniques, this study aims to bridge the existing gaps and provides the following contributions.

- We consider a wide variety of features spanning economic considerations, environmental factors, energy use, electricity consumption, and renewable energy dimensions, ensuring a comprehensive representation of potential determinants of CO₂ emissions.
- We analyze data for 20 high-HDI countries, covering a span of 25 years, providing a significant temporal and spatial scope for the study. The data are exported from a reputable online source, ensuring the reliability and quality of the dataset used for analysis. The study of CO₂ emissions within high-HDI countries offers not just an insight into the emission themselves but also into the broader narrative of environmental responsibility and sustainability. It is within these nations that the battle against climate change can be most effectively strategized and executed.
- We employ five different statistical models utilizing Ordinary Least Squares (OLS), fixed effects, and random effects based models on the panel dataset. Panel data combine elements of both cross-sectional data and time series data which track multiple subjects such as countries across different points in time. This varied approach in modeling has allowed us to conduct a comprehensive comparative analysis, crucial for pinpointing the key predictive features. Additionally, the diversity of these models enhanced our ability to effectively analyze cross-sectional time series data, providing a deeper understanding of temporal trends and variations across different countries.
- We adopt a structured, two-phase approach. In the initial phase, we focus on identifying the significance of different variables that influence CO₂ emissions. The subsequent phase aims at forecasting CO₂ emissions for the forthcoming three years and groups countries into clusters according to their emission trajectories. This clustering plays a pivotal role in deciphering and classifying the diverse emission trends across nations, offering critical insights that can inform the creation of precise environmental policies and foster international collaboration.

An illustration of our comprehensive two-stage framework is presented in Figure 1. The numerical study has successfully validated our forecasting models, achieving a higher accuracy in predicting CO₂ emissions and closely mirroring actual trends. This highlights the proposed framework's effectiveness in providing dependable emission projections for the near future.

The rest of the paper is organized as follows. Section 2 provides a comprehensive review of the literature. Section 3 presents the panel dataset used in this study. Section 4 describes the preliminaries, definitions, and steps of our proposed method. Section 5 highlights the simulations, computational results, and effectiveness of our method. Finally, Section 6 provides concluding remarks and suggestions for future research.

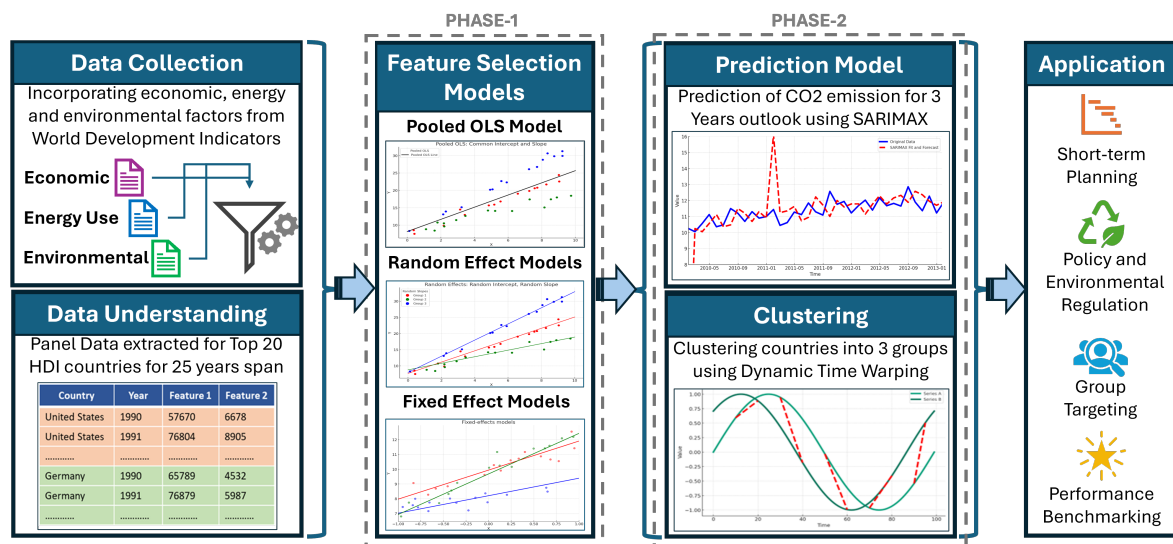


Figure 1. An overview of the utilized framework.

2. Literature Review

The increasing focus on carbon emissions prediction is a response to the urgent need to address climate change and its impacts [53]. Figure 2 illustrates a marked rise in carbon emissions prediction research, showing an upswing in publications from 2010 onwards. This trend underscores an increasing academic and scientific interest in comprehending and forecasting carbon emissions, driven by escalating concerns over global climate change. As the primary driver of global warming, CO₂ emissions are a key focus in efforts to mitigate climate change [6]. Accurate predictions are essential for informing and guiding policy decisions, developing effective strategies for emission reduction, and meeting international climate targets, such as those outlined in the Paris Agreement [54]. Additionally, these studies play a vital role in advancing our understanding of the complex interplay between economic development, energy use, technological advancement, and environmental sustainability. By improving the accuracy and reliability of emission forecasts, this growing body of research contributes significantly to global efforts to protect the environment and ensure a sustainable future [55]. The extant literature on carbon emissions prediction, comprising investigations that pinpoint critical determinants and guide policy formulation, can be broadly categorized into two distinct streams: one consists of research dedicated to a single country or various regions within a single nation, while the other encompasses studies that concurrently address carbon emissions across multiple countries.

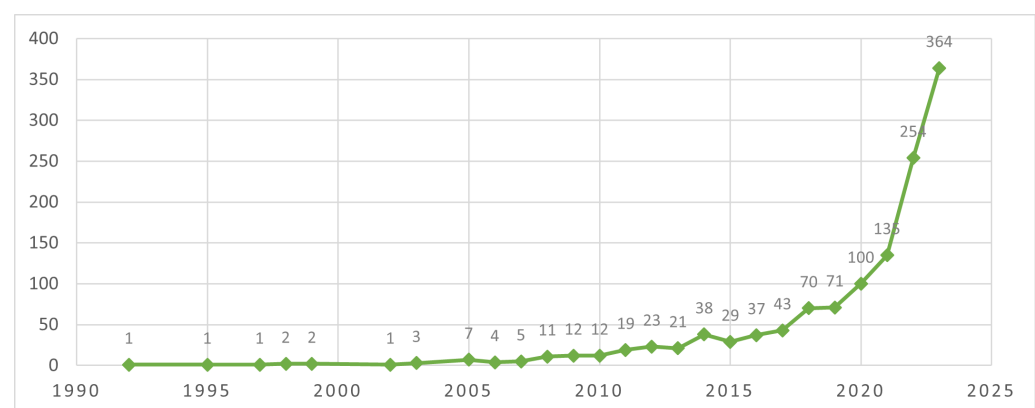


Figure 2. Trend of research in carbon emissions prediction from analysis of 1266 documents.

The United States is a leader in climate research, supported by top-tier universities and institutions specializing in carbon emissions prediction, playing a prominent role in efforts to understand and mitigate climate change [56,57]. China's focus on carbon management research reflects its significant economic growth and population, underscoring its crucial role in global environmental sustainability [58,59]. Germany is recognized for its dedication to renewable energy and climate technology, contributing significantly to emission prediction research [60]. The United Kingdom, with extensive research centers, advances climate science and policy [61]. Australia's research addresses unique environmental challenges and contributes to national and international carbon management strategies [62]. Canada's emphasis on sustainable development integrates environmental, economic, and social factors into emission prediction research [63]. Japan leverages its technology sector to address its vulnerability to climate change, focusing on technology, policy, and environmental science [64]. Post-Paris Agreement, France leads in developing innovative carbon reduction strategies [65]. India's research addresses the challenges and opportunities presented by rapid development and its large population [66]. Scandinavian countries, known for their commitment to sustainability, serve as models with insightful research on emissions [67]. Beyond country-specific studies, the literature includes analyses of multiple countries, such as the G20, where studies have forecasted CO₂ emissions and examined influential factors [68–70]. The OECD (Organization for Economic Cooperation), a larger group of countries, is another area of focus for carbon emissions research [71–73]. Additionally, studies on European, Middle-Eastern, and Asian countries further enrich the field [46,74,75]. Researchers often group countries by criteria such as geography or political relations, adding depth to the comparative understanding of emissions.

Clustering countries based on their CO₂ emissions profiles is pivotal for tailored environmental policies and actions. The spatial clustering of industrial CO₂ emissions efficiency across China's provinces highlights the need for targeted regional policies [76]. Such clustering methodologies dissect the intricate relationship between CO₂ emissions and economic growth, enabling tailored interventions suited to diverse developmental trajectories [77]. For instance, the Kyoto Protocol's lack of consideration for agriculture and forestry underscores the value of clustering for more specific policies [78]. Similarly, the European Union's Emission Trading System exemplifies how market-based mechanisms can lead to significant greenhouse gas reductions, showcasing the importance of strategic and targeted policies [79]. Analysis of CO₂ emissions determinants in Far East countries emphasizes the need for green urbanization, industrial regulation, and renewable energy adoption, illustrating how clustering can guide urgent regional actions [80]. Insights from structural decomposition analysis reveal varying impacts across country groups, necessitating a multifaceted approach to global warming that acknowledges diverse emission drivers [4]. By categorizing countries into clusters reflecting unique emission characteristics and challenges, policymakers can devise more effective, context-specific strategies for CO₂ control.

Building upon the existing literature on carbon emissions prediction, extensive research has been conducted to identify and evaluate the effectiveness of various indicators influencing carbon emissions. These indicators, ranging from economic factors such as GDP and industrial activity to energy consumption metrics and technological advancements, are paramount in constructing accurate predictive models [81]. Economic indicators have been extensively studied to understand their relationship with carbon emissions. Gross Domestic Product (GDP) has been a focal point, with numerous studies exploring the Environmental Kuznets Curve (EKC) hypothesis, which postulates an inverted U-shaped relationship between per capita income and environmental degradation [82]. However, recent critiques highlight that the EKC oversimplifies the relationship between economic

growth and environmental degradation [83]. It often neglects the externalization of emissions through global trade, where high-income countries outsource pollution-intensive activities to low-income nations, and assumes uniform development trajectories, overlooking global inequalities [84]. Industrial activity, another crucial economic indicator, has been linked with higher carbon emissions due to the energy-intensive nature of manufacturing processes [85]. Energy consumption, particularly the reliance on fossil fuels, has been consistently identified as a primary indicator of carbon emissions. The shift towards renewable energy sources is reflected in reduced emission intensity in various predictive studies [86]. Technological innovation, measured through indicators such as energy efficiency improvements and cleaner technologies, has also been shown to mitigate emissions [87]. Societal behaviors and policy indicators, including urbanization rates, population density, and environmental regulations, influence energy demand and consumption patterns, thereby affecting carbon emissions [88]. Additionally, carbon pricing mechanisms, such as carbon taxes and emission trading schemes, have been analyzed as indicators for incentivizing emission reductions [89]. Identifying these indicators is critical for policymakers to target influential factors that can lead to substantial emission reductions. Predictive models incorporating these indicators provide a basis for scenario analysis, enabling decision-makers to forecast outcomes of various policy interventions and strategize effectively for climate change mitigation [90]. The identification and analysis of indicators in carbon emissions prediction remain essential components of climate change research, offering quantitative insights into emission dynamics and strategies to meet global climate goals. Ongoing research in this domain continues to refine these indicators and improve the predictive power of models, contributing to more informed and effective environmental policymaking.

In exploring CO₂ emissions predictions, a range of methodologies has been applied, each revealing distinct advantages and inherent limitations. In one study, the author utilizes Artificial Neural Networks (ANNs) and Vector Autoregression (VAR) for forecasting emissions in the USA, shedding light on the impact of economic growth and energy consumption on CO₂ levels. However, the study concedes its inability to accurately predict output spike forms, suggesting potential inaccuracies during abrupt economic or environmental shifts [91]. In another study, authors adopted System Dynamics (SD) with STELLA's simulation for city-level CO₂ forecasting in Japan, achieving high accuracy. Yet, their model's assumption of a relatively constant urban structure might not fully capture the dynamic and evolving urban landscapes [92]. Regression analyses have been employed to investigate GHG emission trends across OECD countries, noting the role of energy prices and economic output. They acknowledge the possibility of specification errors and the difficulty in capturing the complex drivers of emissions across various national contexts [71]. Another study investigated the influential factors of carbon emission intensity using symbolic regression, pinpointing GDP, population, and urbanization as critical factors. Nonetheless, the study's emphasis on factor occurrence over precise impact restricts a deeper understanding of causality and influence magnitude on CO₂ emissions [72]. These efforts collectively highlight the intricate challenges in predicting CO₂ emissions, from specific model limitations to broader data availability issues and the dynamic interplay of socio-economic and environmental factors. To overcome some of the limitations of regression-based methods, advanced predictive models such as ARIMA and ANN have been employed [93,94]. ARIMA models, widely used for time series forecasting, offer simplicity and interpretability, making them suitable for emissions prediction in stable environments with clear trends and seasonality [95]. However, their reliance on linear assumptions and limitations in handling non-linear relationships or abrupt changes restrict their applicability in more dynamic contexts [81]. In contrast, ANN models excel in capturing non-linear and complex interactions between variables, making them highly

effective for datasets with intricate emission patterns [96]. Despite their robustness, ANN models require large datasets for training and are prone to overfitting, particularly in data-scarce scenarios [87].

Regarding CO₂ emissions clustering, diverse methodologies showcase their unique constraints. One study employs DEA window analysis and Moran's I statistic for spatial clustering in China's provinces [76], while another leverages Symbolic Regression for Regression Discontinuity (SRRD) alongside the Apriori algorithm, clustering countries by their CO₂ emissions and economic growth models, thus innovatively identifying underlying models [77]. Further, Structural Decomposition Analysis (SDA) is used to dissect factors driving global GHG emission growth, providing insights at both country and industry levels [4]. SDA is particularly valuable for identifying policy-relevant drivers of emissions, such as economic activity and fuel mix, but its reliance on input–output tables can limit its timeliness and adaptability to dynamic changes [89]. Despite these advancements in clustering methodologies, the integration of both prediction and clustering approaches remains unexplored, presenting a gap in the literature that merits further investigation for a more comprehensive understanding of CO₂ emissions dynamics.

Tables 1 and 2 provide a comprehensive overview of various research studies focused on CO₂ emissions analysis, categorizing them based on the methodologies employed, their key findings, and the application of different modeling techniques. Specifically, it delineates whether each study incorporated supervised learning (SL), utilized panel data (PD), considered unsupervised learning (UL), and analyzed multiple countries (MCs). This classification helps in identifying methodological trends within the existing literature. While numerous studies have applied either supervised learning (SL), panel data (PD), unsupervised learning (UL), or multiple countries (MCs) analysis independently, our approach combines these elements to provide a more holistic and comprehensive analysis of CO₂ emissions. This integrative approach is designed to leverage the strengths of both supervised and unsupervised learning, while the use of panel data from multiple countries will enhance the generalizability and applicability of our findings across different geographical and economic contexts.

Table 1. Summary of methods and findings in carbon emissions prediction research (Part 1).

Reference	Method	Finding	SL/PD/UL/MC	Features
[91]	Artificial Neural Network and Vector Autoregressive Estimator	Importance of renewable energy consumption on CO ₂ emissions	Y/N/N/N	Economic growth, net trade, renewable energy types
[97]	Mixed Data Sampling Regression and Back Propagation Neural Network	Accuracy improvement using hybrid forecasting model	Y/N/N/N	Quarterly economic growth data to predict annual CO ₂ emissions
[56]	Structural Augmented Dynamic Factor Model	Model effectively forecasts and nowcasts U.S. CO ₂ emissions	Y/N/N/N	Macroeconomic variables, specifically industrial production indices and residential utilities
[57]	Extended Logarithmic Mean Divisia Index Decomposition	Main factor increasing US emission is the scale effect, while the technology effect significantly reduces emission	Y/N/N/N	Income, population, energy intensity, emission coefficient, economic structure, and energy consumption structure

Table 1. Cont.

Reference	Method	Finding	SL/PD/UL/MC	Features
[98]	Gray Wolf Optimizer combined with a Support Vector Machine	New model effectively predicts carbon emissions, showing better performance compared to traditional models	Y/N/N/N	Economic growth, population, energy consumption, industrial structure, urbanization rate, and technological advancement
[58]	STIRPAT Model with Ridge Regression and Scenario Analysis	Differential peak attainment of carbon emissions among 16 cities	Y/Y/N/N	A total of 16 socio-economic factors including GDP, population, industry output values, and energy consumption
[60]	Linear Regression for marginal emission factors, and a Multilayer Perceptron	Smart charging reduces CO ₂ emissions by 1% to 10%, by shifting to times with lower marginal factors	Y/N/N/N	Historical load data, temperature data, calendar information, and possibly other not specified variables for forecasting
[99]	AIM/Backcasting Model	Achieving an 80% reduction in CO ₂ emissions by 2050 is feasible with early actions	Y/N/N/N	Consumer preferences, technological options including current and future best available technologies, socio-economic conditions
[92]	System Dynamics using STELLA's simulation engine	High forecasting accuracy with MAPEs less than 20% for all activities	Y/N/N/N	Socio-economic data, residential statistics, service and industrial output, transportation data
[65]	Backpropagation Neural Networks	CO ₂ emissions are predicted to increase by 26.5–36.5% by 2030	Y/N/N/Y	Economic growth, energy consumption, share of renewable energy
[68]	Quantile Regression	Negative marginal relationship between CO ₂ emissions and both socio-economic indicators	Y/Y/N/Y	GDP per capita, fossil fuel consumption, urbanization, trade openness, population density, LPI, and HDI
[69]	Adaptive Neuro-Fuzzy Inference System	Model efficiently predicted CO ₂ emissions with high accuracy	Y/Y/N/Y	Energy consumption and economic growth
[70]	Sparrow Search Algorithm, Fractional Accumulation Grey Model, and Support Vector Regression	Significantly improves prediction accuracy of carbon emissions	Y/Y/N/Y	Economic, social, technological, energy, and trade factors
[44]	Slack-Based Measure Data Envelopment Analysis and Malmquist–Luenberger Index	Renewable energy consumption increases average energy efficiency but decreases energy productivity	Y/Y/N/Y	Labor, gross fixed capital formation, non-renewable energy consumption, renewable energy consumption, GDP, and CO ₂ emissions

Table 2. Summary of methods and findings in carbon emissions prediction research (Part 2).

Reference	Method	Finding	SL/PD/UL/MC	Features
[71]	Series of regression analyses	Energy prices, economic output, and environmental governance are key in explaining GHG emission variations	Y/Y/N/Y	Climate, population pressure, economic output, technological development, industrial structure, energy prices, environmental governance, pollution abatement and control expenditures, environmental pricing
[79]	Linear Model and Natural Logarithm	A total of 24 countries managed to reduce both CO ₂ and GHG emission	N/N/Y/Y	CO ₂ and GHG emissions
[80]	Panel data-fixed effect model, incorporating time trends	Urbanization, economic growth, and trade openness significantly determine CO ₂ emissions	Y/Y/N/Y	Urbanization, GDP, trade openness
[4]	Structural Decomposition Analysis	Rapid global economic growth is a dominating driving force	N/N/Y/Y	Industry energy use, CO ₂ emissions and emission to air in the world
[72]	Symbolic Regression	Influential factors vary by country, with GDP being the most frequent	Y/Y/N/Y	GDP, total population, industrialization, urbanization, technological innovation, and foreign direct investment
[73]	Dynamic Panel Autoregressive Distribution Lag model	Urbanization has a negative impact on CO ₂ emissions per capita, total CO ₂ emissions, and CO ₂ emissions intensity in OECD countries	Y/Y/N/Y	Urbanization rate, GDP per capita, energy intensity, industry, transportation, residential sectors, population
[76]	The DEA window analysis approach and Moran's Index	The northwest region has the highest CO ₂ emissions reduction potential in the industrial sector	N/Y/Y/N	Labor, capital, energy consumption, value-added of industrial enterprises, and CO ₂ emissions
[78]	Statistical Analysis and Clustering	Kyoto Protocol has not considered its effects on agricultural and forestry	N/Y/Y/Y	Greenhouse gas emission data from the agricultural and forestry sectors
Our Proposed Approach	Five statistical models, SARIMAX, and DTW	Identifying important factors, 3 years CO ₂ emissions prediction into the future, clustering 20 countries in 3 groups	Y/Y/Y/Y	Mentioned in Table 3

Table 3. Indicators of environmental impact and energy use.

Indicator	Definition	Significance
CO ₂ emissions (kt) (Co)	This indicates the emissions coming from the burning of fossil fuels and the manufacture of cement. Carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring are included here.	Direct indicator of a country's contribution to global CO ₂ levels and climate change.
CO ₂ emissions (metric tons per capita)	Carbon dioxide emissions in metric tons per capita as defined above.	Reflects the average CO ₂ footprint of individuals, useful for comparing among populations.
Energy use (kg of oil equivalent per capita) (EU)	Energy use calculated as primary energy before transformation to other end-use fuels, which is equal to indigenous production plus imports and stock changes, minus exports and fuels supplied to ships and aircraft engaged in international transport.	Higher per capita energy use typically correlates with higher CO ₂ emissions.
Renewable energy consumption (% of total final energy consumption) (RE)	The share of renewable energy in total final energy consumption.	The greater the reliance on renewables, the lower the CO ₂ emissions from energy consumption.
Electricity production from oil, gas, and coal sources (% of total) (EP)	Electricity production sources are the fuels used to generate electricity. Oil encompasses crude oil and petroleum products. Gas refers to natural gas but excludes natural gas liquids. Coal includes all coal and brown coal, both primary (including hard coal and lignite—brown coal) and derived fuels (including patent fuel, coke oven coke, gas coke, coke oven gas, and blast furnace gas). Peat is also considered in this category.	Indicates the dependency on fossil fuels for electricity, which is directly related to CO ₂ emissions and environmental impact.
Forest area (sq. km) (F)	This considers land under natural or planted stands of trees of at least 5 meters in situ, regardless these are productive or not. However, this does not consider tree stands in agricultural production systems (i.e., in fruit plantations and agroforestry systems) and trees in urban parks and gardens.	Forests are carbon sinks; larger forest areas can mitigate CO ₂ emissions by absorbing more CO ₂ .
Forest area (% of land area) (Fa)	Proportion of land covered by forests as defined above.	Higher percentages indicate greater potential for natural CO ₂ absorption and climate regulation.
Electric power consumption (kWh per capita) (EPC)	It measures the production of power plants and combined heat and power plants less transmission, distribution, and transformation losses and own use by heat and power plants.	High per capita consumption often leads to higher CO ₂ emissions if electricity is produced from fossil fuels.
GDP per unit of energy use (PPP \$ per kg of oil equivalent) (G)	This represents the PPP GDP per kilogram of oil equivalent of energy use. PPP GDP is gross domestic product converted to current international dollars using purchasing power parity rates based on the 2017 ICP round. An international dollar has the same purchasing power over GDP as a U.S. dollar has in the United States.	Efficient use of energy for economic output can suggest a lower CO ₂ emissions intensity.
Total greenhouse gas emission (kt of CO ₂ equivalent) (TG)	This is composed of CO ₂ totals omitting short-cycle biomass burning (such as agricultural waste burning and savanna burning) but incorporating other biomass burning (such as forest fires, post-burn decay, peat fires, and decay of drained peatlands), all anthropogenic CH ₄ sources, N ₂ O sources, and F-gases (HFCs, PFCs, and SF ₆).	Provides a broader picture of a country's impact on warming, including but not limited to CO ₂ emissions.

3. Data Understanding

For the development of a global CO₂ emissions prediction model, we have meticulously sourced our data from the World Development Indicators database, a comprehensive repository maintained by the World Bank [100]. This dataset provides an extensive range of economic, social, and environmental indicators, which have been instrumental in understanding and forecasting global CO₂ emissions trends. Table 3 summarizes the information about the features and their importance in the modeling process.

We have attempted to visualize the data to gain a deeper understanding of the underlying patterns and correlations within the data to assist the development of the model. It is possible to identify trends and uncover underlying structures by employing a variety of visualization techniques. Firstly, we examine the trend in CO₂ emissions among various countries. The available data has been compared from two distinct periods: 20 years preceding 2010 and 5 years following 2010. Understanding emission patterns over time and identifying potential shifts in environmental policy or industrial practices are crucial for understanding how emission patterns have evolved. Based on Figure 3, countries exhibit three distinct emission trends. The first group, exemplified by Singapore and Norway, exhibits an upward trend in CO₂ emissions following 2010. As a result of this increase, emission may have increased either due to an increase in industrial activity or energy consumption. In contrast, the second group, including countries such as the United States and Denmark, exhibits a decrease in emission following 2010. As a result of successful environmental policies, the shift towards renewable energy sources, or improvements in energy efficiency, this reduction may occur. Additionally, for countries such as Slovenia and Austria, our analysis revealed a consistent trend in CO₂ emissions without any significant decrease or reduction observed over the two periods.

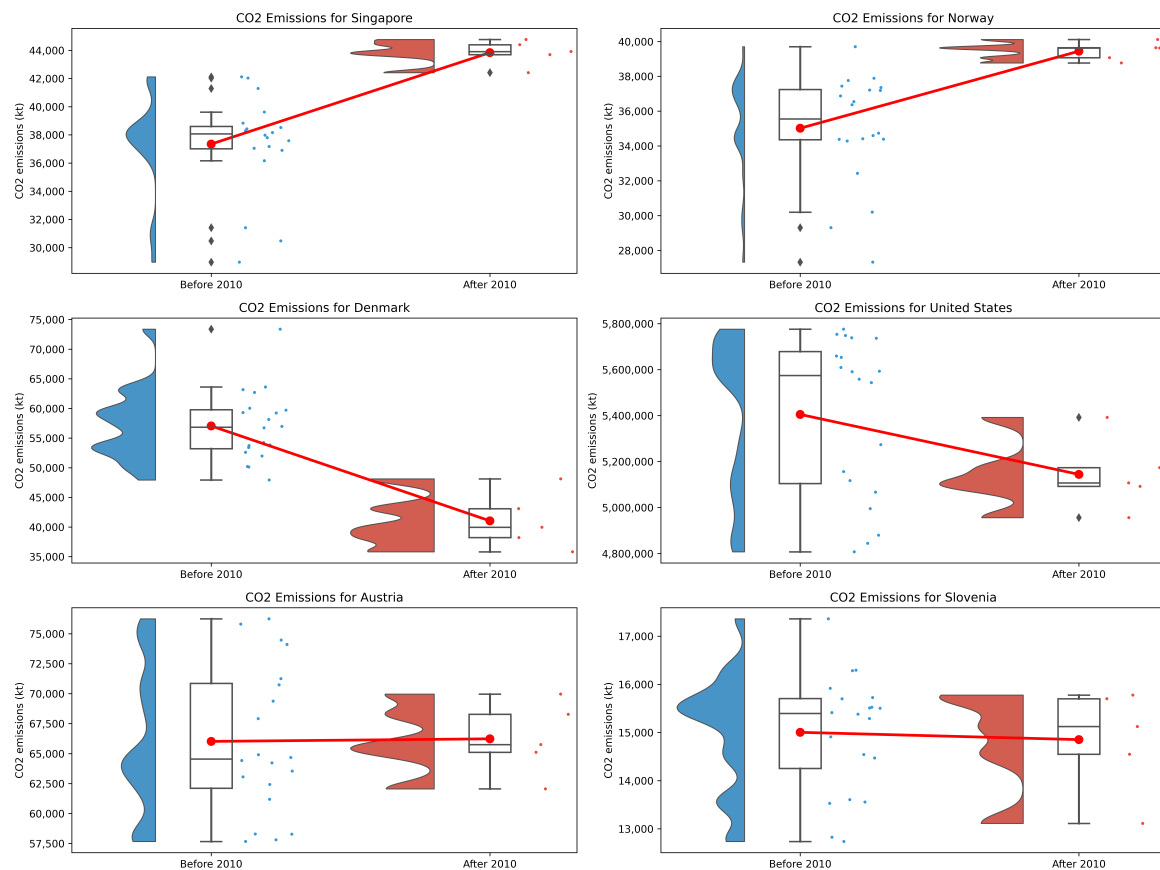


Figure 3. Different carbon emissions patterns over years in countries.

To get an overall idea about the pattern of CO₂ emissions, it is necessary to categorize countries based on their long-term emission trends. Analyzing trends in CO₂ emissions provides initial insights. However, the historical emission patterns of each country should be examined. As shown in Figure 4, a hierarchical clustering approach coupled with Ward's method [101] demonstrates a dendrogram that provides a systematic method for grouping countries by their normalized CO₂ emissions. Although sensitive to outliers, the Ward method ensures that all the countries grouped together exhibit similar emission dynamics.

by minimizing the total within-cluster variance. The height of the merge lines indicates distances between clusters, revealing three distinct groups for the countries.

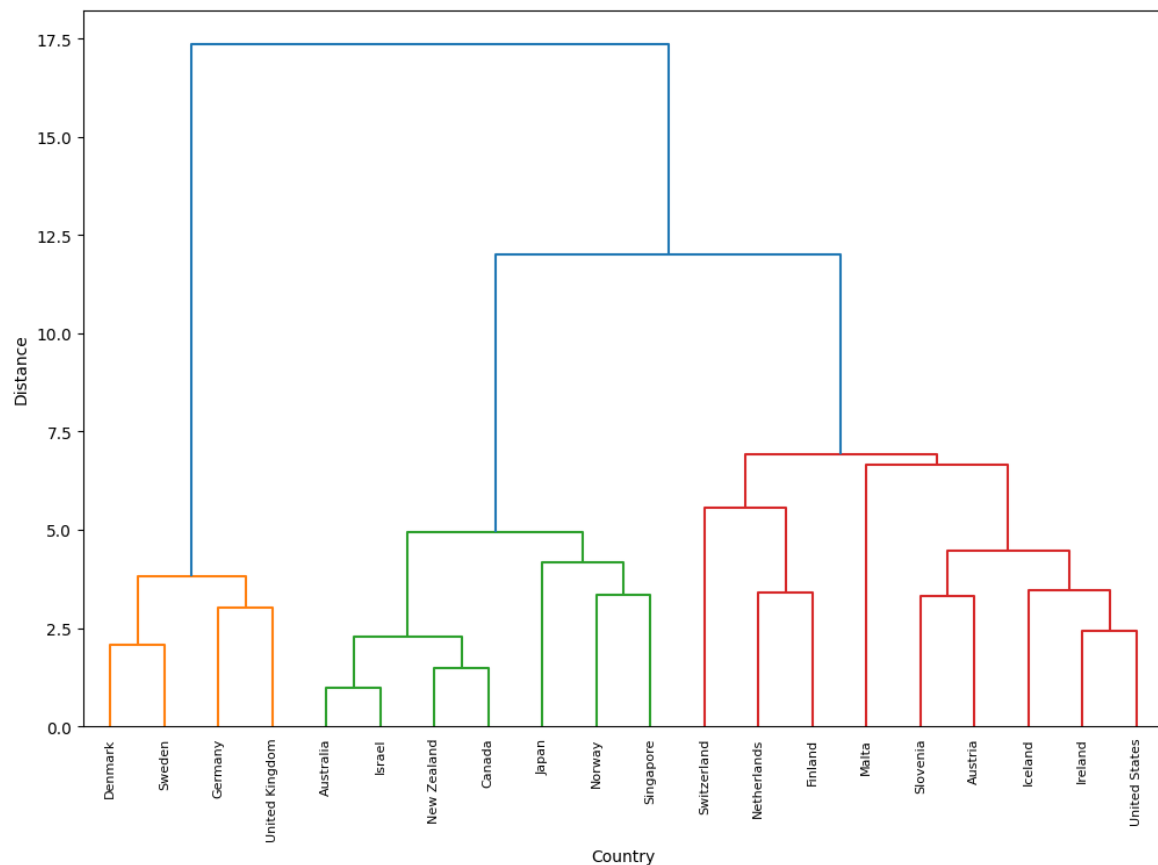


Figure 4. Dendrogram visualization of grouped countries by their normalized CO₂ emissions. Each color represents a distinct group of countries.

To analyze the features and their correlations, Figure 5 is presented. Figure 5 provides a correlation matrix, the size and color of the circles indicate the strength and direction of the correlation between the variables. These relationships may shift over time or across regions due to policy and economic changes. In this plot, the larger circles represent stronger correlations, whereas blue represents positive correlations, and red indicates negative correlations. It is noteworthy that ‘Renewable energy consumption’ and ‘Electric power consumption’ exhibit a highly positive correlation (quantified at 0.82). Also, there is a notable positive correlation between ‘Electric power consumption’ and ‘Energy use’, evidenced by a correlation coefficient of 0.88. Conversely, a strong negative correlation of -0.83 is observed between ‘Electricity production from oil, gas, and coal sources’ and ‘Renewable energy consumption’. This inverse relationship suggests that as one variable increases, the other tends to decrease, possibly reflecting a shift towards renewable energy sources or improvements in production efficiency. These correlations have a considerable impact on our modeling approach. When both features, highly correlated, emerge as significant predictors, it is more suitable to use only one of them to avoid multicollinearity. Multicollinearity can lead to skewed predictions and hinder the interpretability of the model, thus careful selection of predictors is crucial to maintain the accuracy and reliability of our analysis. To further understand the distribution and evolution of the indicators over time, Figure 6 is presented.

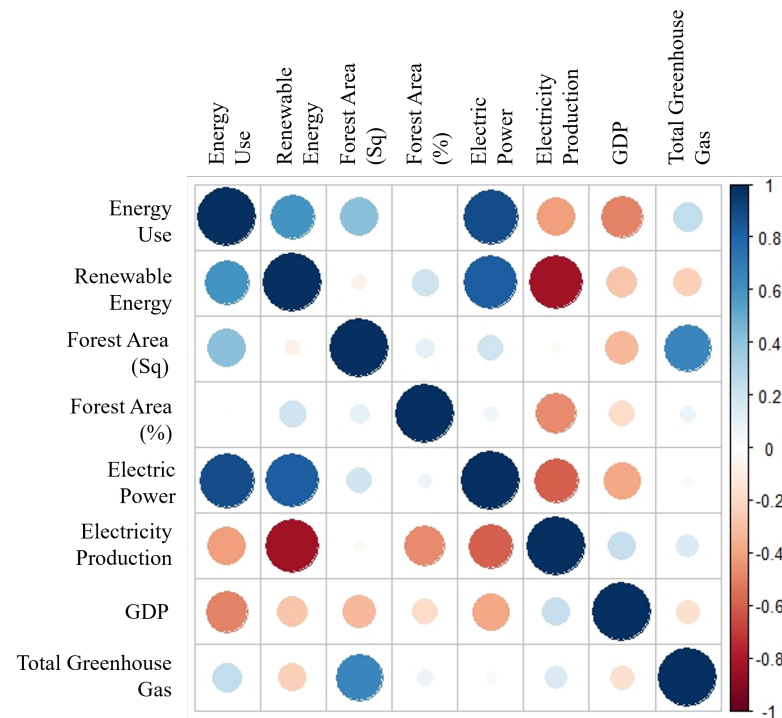


Figure 5. Heatmap visualization of features.

Using ridge plots in Figure 6, we present insight into four characteristics of our panel dataset based on their distributional characteristics. Figure 6a illustrates the consumption of electricity over time. The layer represents a specific year, and the width indicates the density of countries consuming specific amounts of electric power. More countries have consumed larger quantities of electricity in the last years, as indicated by the widening of the plot. Figure 6b demonstrates an increasing variability among countries in energy use efficiency from 1990 to 2014. Figure 6c presents considerable variation in the height of layers for different years and signifies a general increase in energy consumption, though at varying rates among different countries. This variation underscores the diverse energy consumption trajectories globally. In Figure 6d, the depiction of changes in forest area over the 25 years is marked by distinct annual layers. Notably, the peak of these layers maintains a relatively stable position, without significant shifts to the right or left. This stability suggests that there hasn't been a dramatic change in forest areas globally over the observed period. Next, parallel coordinates plots are demonstrated to provide a multi-dimensional perspective on how various factors relate to CO₂ emissions in Figure 7. Across these visualizations, each line corresponds to a country, while the position along each axis represents a country's normalized value for the features.

According to Figure 7a, higher levels of energy consumption do not necessarily lead to higher CO₂ emissions across countries, suggesting the importance of other factors. Similarly, Figure 7b does not indicate a clear correlation between 'Forest area (%)', 'Renewable energy', and CO₂ emissions, meaning that there exist other significant contributions such as industrial activity levels or environmental regulations, as evidenced by intersecting paths between these axes. Based on the Figure 7c, there is no direct relationship between 'Electric power', 'Electricity production from oil, gas, and coal sources', and CO₂ emissions, either. However, in Figure 7d, the pattern is more predictable. In this plot, a lower 'GDP per unit of energy use' correlates with higher greenhouse gas emissions, leading to higher CO₂ emissions. These visualizations, while insightful, highlight the challenges in identifying the most influential factors for predicting CO₂ emissions. Therefore, a comprehensive feature selection process, coupled with advanced modeling techniques for panel data, is essential to

accurately pinpoint the key drivers of CO₂ emissions. This approach is vital for developing effective models that can navigate the complex factors influencing CO₂ emissions.

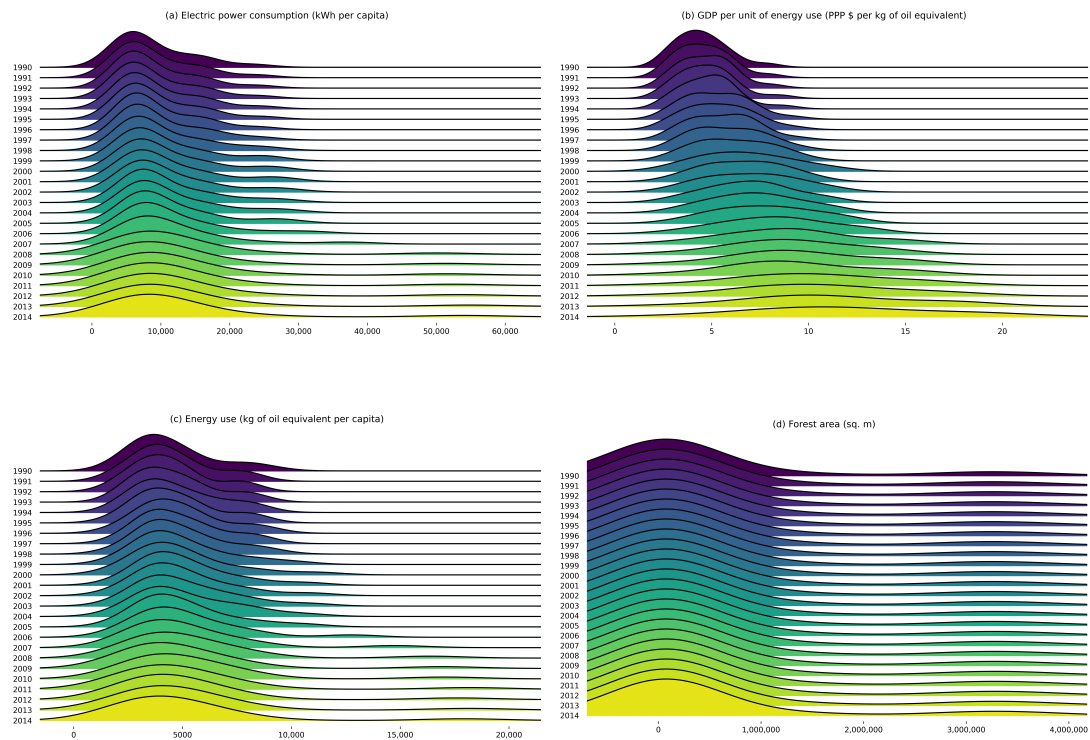


Figure 6. Ridge plots of the four indicators showing the distribution and evolution of them over time.

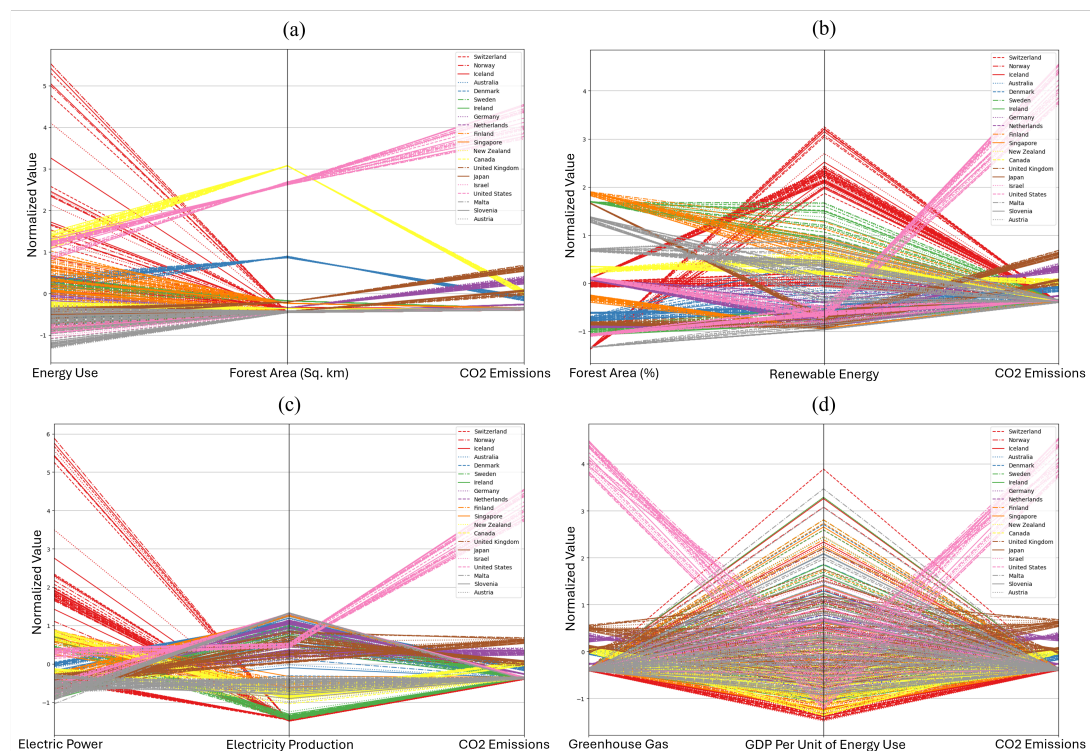


Figure 7. Parallel coordinates plots of the features to provide a multi-dimensional perspective.

4. Methodology

In this study, we employed a two-phase analytical approach to analyze CO₂ emissions. Initially, five statistical models including OLS, fixed effects, and random effects are used to

identify key features affecting CO₂ emissions. Subsequently, we implement both supervised (SARIMAX) and unsupervised (DTW) learning models for a thorough examination of emission trends.

Our methodological advancements address limitations of traditional linear regression models in handling panel datasets. By using models that account for unobserved heterogeneity across countries and time, we have achieved more robust inferences, crucial given the diversity and dynamic nature of high-HDI countries' emissions. SARIMAX's (which stands for Seasonal AutoRegressive Integrated Moving Average with eXogenous variables) ability to integrate seasonal patterns and external variables offers a nuanced understanding of emission, while DTW's refined clustering technique provides insights into countries' emission trajectories, highlighting convergences and divergences in their patterns over time. This comprehensive methodology facilitates a deeper understanding of the factors driving CO₂ emissions, enhancing the potential for effective policy formulation.

4.1. Terminology

4.1.1. Pooled OLS Model

Pooled Ordinary Least Squares (Pooled OLS) is a prevalent econometric technique for analyzing panel data, which involves multiple observations over time for a number of cross-sectional units. This method is particularly useful for datasets where the focus is on understanding the collective impact of variables across different entities, such as countries or firms, over time. Pooled OLS treats the panel data as a large cross-sectional dataset, essentially aggregating all time periods and entities, which simplifies the analysis but also imposes certain limitations. The core idea of Pooled OLS is to estimate a single regression model for the entire panel dataset. In Pooled OLS, the regression equation is formulated as follows:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \epsilon_{it} \quad (1)$$

In the above equation, Y_{it} is the dependent variable for the i -th cross-sectional unit (e.g., country) at time t , X_{it} represents independent variables, β_0 and β_1 are coefficients, and ϵ_{it} is the error term. This model assumes that the slope coefficients (β) are constant across all cross-sectional units and over time. The estimation of the model parameters is typically achieved through the OLS method, which involves minimizing the sum of squared differences between the observed values and the values predicted by the model. One of the primary limitations of Pooled OLS is its assumption of homogeneity, implying that the slope coefficients are constant across all entities and over time. This assumption often leads to biased estimates in the presence of unobserved heterogeneity among cross-sectional units. Moreover, Pooled OLS assumes no autocorrelation and homoscedasticity in error terms, which is often violated in panel data settings. Recent studies emphasize the need for careful application of Pooled OLS, especially in the presence of cross-sectional dependence and heterogeneity.

4.1.2. Random Effects Model

The Random Effects (RE) model stands as a pivotal approach in the realm of panel data analysis, particularly when addressing data that comprises multiple observations across various entities over time. Distinguished from the Pooled OLS approach, the RE model acknowledges the potential heterogeneity across cross-sectional units (such as countries, companies, or individuals) by incorporating a unique error component for each entity. This model is especially pertinent when the unobserved entity-specific effects are assumed to be uncorrelated with the explanatory variables, a condition that makes it distinct from the Fixed Effects (FE) model. The RE model is predicated on the notion that while all entities in the panel share common behavior captured by the overall intercept and slope coefficients,

there is also an idiosyncratic, entity-specific random component. The model is typically represented as follows:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + u_i + \epsilon_{it} \quad (2)$$

Here, Y_{it} denotes the dependent variable for entity i at time t , X_{it} represents the independent variables, β are the coefficients to be estimated, u_i is the unobserved random effect unique to each entity, and ϵ_{it} is the idiosyncratic error term. The estimation of the RE model usually involves Generalized Least Squares (GLS), a method that efficiently deals with the panel-specific variance component u_i . The critical limitation of the RE model lies in its assumption of no correlation between the entity-specific effects and the independent variables. If this assumption is violated, the RE model may yield biased and inconsistent estimates. Recent methodological advancements in panel data analysis have focused on developing more robust random effects models that can handle situations where traditional RE assumptions might not hold.

4.1.3. Fixed Effects Model

The FE model is a cornerstone in panel data analysis, particularly valuable when dealing with data that exhibit variability across entities and over time. This model is particularly adept at analyzing effects within entities (such as countries, firms, or individuals) by controlling for entity-specific characteristics. The FE model is distinguished by its ability to account for unobserved heterogeneity when this heterogeneity is correlated with the independent variables in the model, a key point that differentiates it from the RE model. The FE model operates under the premise that there are unique attributes inherent to each entity that may influence the dependent variable. These attributes, while potentially correlated with the independent variables, are assumed to be constant over time. The model can be represented as follows:

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \quad (3)$$

Here, Y_{it} represents the dependent variable for entity i at time t , X_{it} are the independent variables, β represents the coefficients associated with each independent variable, which are to be estimated, α_i symbolizes the unobserved individual-specific effect, and ϵ_{it} is the idiosyncratic error term. FE models are typically estimated using the Least Squares Dummy Variable (LSDV) approach or the within transformation. The LSDV approach involves including a dummy variable for each entity, whereas the within transformation, more commonly used due to its efficiency, removes the entity-specific effects by demeaning the variables by their means within each entity. This transformation focuses on the deviations from each entity's mean, eliminating the entity-specific effects and isolating the impact of the independent variables.

4.1.4. Dynamic Time Warping

In this research, Dynamic Time Warping (DTW) is employed as a key methodology to cluster countries based on the trends in their CO₂ emissions. This approach is particularly suited for analyzing panel datasets where temporal patterns in CO₂ emissions across countries may not align perfectly in time. DTW's ability to measure similarities between temporal sequences with varying lengths and speeds makes it an ideal tool for grouping countries with similar emission trends, despite discrepancies in timing or rate of change. DTW is used to compare the temporal sequences of CO₂ emission between different countries. Each country's emission data form a time series, $X = \{x_1, x_2, \dots, x_n\}$, and DTW measures the similarity of these time series by aligning them in a manner that minimizes the cumulative distance. This is achieved by constructing a matrix where each element

represents the distance between two points from two different time series (i.e., two different countries' CO₂ emission levels at different times).

4.1.5. SARIMAX

The ARIMA (Autoregressive Integrated Moving Average) model is a cornerstone in time series forecasting, particularly suited for data showing trends and non-seasonal patterns. It is expressed as ARIMA(p, d, q), where p is the order of the autoregressive part, d the degree of differencing, and q the order of the moving average part. The model can be represented as follows:

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - L)^d Y_t = (1 + \sum_{i=1}^q \theta_i L^i) \epsilon_t \quad (4)$$

Here, Y_t is the time series, L is the lag operator, ϕ_i are the parameters of the autoregressive part, θ_i of the moving average part, and ϵ_t is white noise error. While ARIMA is powerful, it has limitations in handling seasonal fluctuations and external influences, which are common in many time series data, including CO₂ emissions. This leads to the evolution into SARIMAX (Seasonal Autoregressive Integrated Moving Average with exogenous variables), an extension of ARIMA that incorporates both seasonality and exogenous factors. SARIMAX is denoted as SARIMAX(p, d, q)(P, D, Q, S), where P, D, Q represent the seasonal elements of the AR, differencing, and MA components, respectively, and S is the periodicity of the seasons.

$$(1 - \sum_{i=1}^p \phi_i L^i)(1 - \sum_{i=1}^P \Phi_i L^{iS})(1 - L)^d(1 - L^S)^D Y_t = (1 + \sum_{i=1}^q \theta_i L^i)(1 + \sum_{i=1}^Q \Theta_i L^{iS}) \epsilon_t + \beta X_t \quad (5)$$

Here, Φ_i and Θ_i are the seasonal AR and MA parameters, D is the order of seasonal differencing, β represents the coefficients of the exogenous variables X_t , and S denotes the seasonality period. SARIMAX offers a substantial advancement over ARIMA in time series forecasting, particularly for complex datasets like CO₂ emissions, due to its enhanced capabilities. It adeptly handles seasonal trends through seasonal differencing and the incorporation of seasonal autoregressive and moving average components, a critical aspect for data exhibiting seasonal variations. Additionally, SARIMAX's ability to include exogenous variables allows for the integration of external factors, enhancing the robustness and comprehensiveness of the model.

4.1.6. MAE

MAE measures the average magnitude of errors in a set of predictions, without considering their direction. It is a linear score, which means all the individual differences are weighted equally. It is calculated as follows:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (6)$$

In the above equation, N is the number of observations, y_i is the actual value of the observation, and $\hat{y}_i$ is the predicted value.

4.1.7. RMSE

RMSE is a standard way to measure the error of a model in predicting quantitative data. It is particularly sensitive to large errors due to its squaring of the residuals. The RMSE is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (7)$$

In the above equation, N is the number of observations, y_i is the actual value of the observation, and $\hat{y}_i$ is the predicted value.

4.1.8. NRMSE

NRMSE is a normalized version of RMSE that facilitates the comparison between models with different scales. It is defined as follows:

$$\text{NRMSE} = \frac{\text{RMSE}}{y_{\max} - y_{\min}} \quad (8)$$

Here, RMSE is the root mean square error, $y_{\max}$ and $y_{\min}$ are the maximum and minimum values of the observed data, respectively. This normalization helps to bring the error metric to a comparable scale, especially useful when dealing with different datasets or units.

The above specific statistical and time series models employed in this study are carefully chosen to analyze carbon emissions across high-HDI countries, considering the complexities of the dataset and research objectives. Pooled OLS, FE, and RE models are utilized to effectively analyze panel data by capturing cross-sectional and temporal variability, with FE controlling for unobserved heterogeneity and RE treating these characteristics as random for robust inferences. The SARIMAX model is selected for its ability to incorporate seasonal patterns and external variables, making it particularly effective for capturing the cyclical and policy-driven dynamics of carbon emissions in time series forecasting. Additionally, DTW is employed to cluster countries based on emission trajectories, even when temporal patterns do not align perfectly, allowing for more intricate comparisons among nations. These models collectively address key limitations in traditional approaches, ensuring a comprehensive and robust examination of emission trends and their determinants, thereby providing policy-relevant insights.

4.2. The Proposed Method

The proposed method presents a systematic approach to analyze CO₂ emissions across multiple countries over time, employing a sequence of econometric models and statistical tests. The overview of the method is shown in Figure 8. The method starts with data collection and preparation, which is essential to ensure the integrity of the subsequent econometric analysis. The initial examination begins with a Pooled OLS model, assuming homogeneity across entities and time. Then, the Breusch–Pagan test is utilized to check whether evidence of heteroscedasticity is found or not. If this statistical test is rejected, then two different RE models are applied (with and without time trend effect). To further refine the model selection, the Wald test is applied to compare the RE models [102]. Following this, the Hausman test is conducted to find out whether FE models are needed or not. Upon this test's indication, we either continue with the RE model or switch to the FE model. We then apply and compare two different FE models (linear regression with country fixed effects and the Generalized Linear Model (GLM) with country fixed effects) using criteria such as the AIC, BIC, and Likelihood to determine the best model in terms of fitting the

data. Based on the best model found in this step, the important features affecting the CO₂ emissions are selected. Next, phase two of the methodology is employed, focusing on an integrated analysis using both supervised and unsupervised learning paths. It provides a thorough and robust examination of CO₂ emissions across countries with different HDI scores. SARIMAX and DTW are applied in this phase using the selected features from phase one. This comprehensive approach not only bolsters the analytical depth but also ensures that our results are robust and can inform impactful policy decisions within the realm of environmental climate change mitigation.

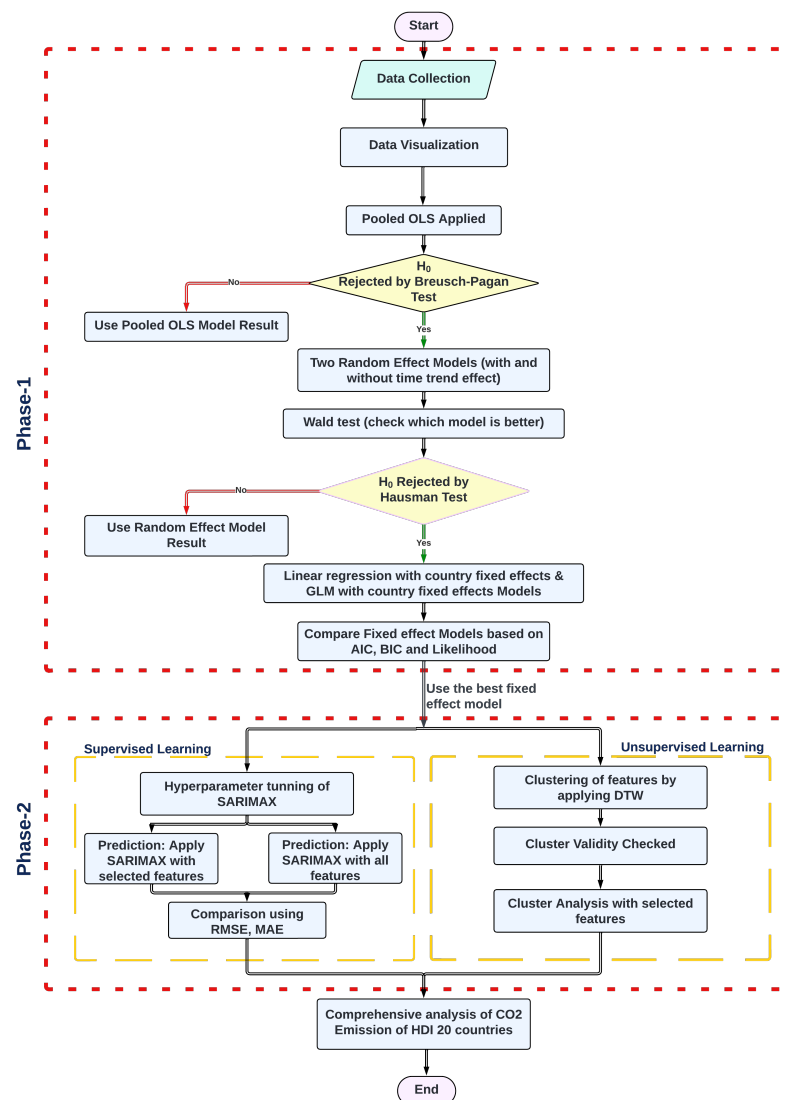


Figure 8. Flow diagram of the proposed method.

5. Results

In this section, we present a detailed analysis of the outcomes derived from the application of outlined methodologies. Initially, we interpret the findings from the five statistical models, followed by a comparative assessment of their performance. Subsequently, leveraging the most effective model identified, we select the most relevant features that significantly influence our analysis. These features are then employed to predict CO₂ emissions over the most recent three-year period of the dataset.

5.1. Phase I: Applying Specialized Models for Panel Dataset

Our analysis comprehensively evaluates various models to ascertain their suitability for the panel dataset. We have explored five distinct models, encompassing a pooled model, two random effect models, and two fixed effect models, to understand their efficacy and limitations. The models' summaries are encapsulated in Table 4, providing insights into their types, objectives, limitations, and potential biases.

Table 4. Summary of the applied models on the panel dataset.

Model	Type	Objective	Limitation	Bias
A	Pooled OLS	Averages the impact of predictors	Cannot account for time-invariant characteristics	Risk of bias from correlated traits
B	Random effects	Averages impacts with group variation	Assumes random effects are normally distributed	Unbiased if effects are normal
C	Random effects with time trend	Considers impacts with group and time trends	Assumes effects are normally distributed	Unbiased if effects are normal
D	Linear regression with country fixed effects	Evaluates impacts, accounting for country specifics	Demands more data	Unbiased, controls for country traits
E	GLM with country fixed effects	Examines relations, incorporating fixed country traits	Assumes normal distribution of dependent variable	Unbiased, addresses country traits

As seen from Table 4, Model A, the pooled model, simplifies the analysis by estimating the average effects of predictor variables without adjusting for time-invariant characteristics, making it suitable for datasets with a smaller number of entities. Model B, the random effects model, offers a balance by assuming these characteristics are random and accounted for them as errors. Model C extends this approach by including a time trend to capture potential systematic changes over time across all groups. Model D employs linear regression with country fixed effects, enhancing the model's ability to control unobserved, time-invariant differences across countries. Lastly, Model E utilizes GLMs with country fixed effects, providing a flexible framework capable of modeling a broader range of data distributions, including non-normal distributions.

5.1.1. Pooled Ordinary Least Squares Model

We initiate our analysis with a Pooled OLS model to assess our dataset comprehensively. The model is structured as follows:

$$C_{oit} = \beta_0 + \beta_1 RE_{it} + \beta_2 Fa_{it} + \beta_3 EPF_{it} + \beta_4 G_{it} + \beta_5 TG_{it} + \beta_6 F_{it} + \beta_7 EU_{it} + \beta_8 EPC_{it} + \epsilon_{it} \quad (9)$$

where C_{oit} denotes the dependent variable for entity i at time t , β values are the coefficients for the respective predictors, and ϵ_{it} represents the error term. The regression output, summarized in Table 5, reveals the significance levels of various predictors on CO₂ emissions.

The analysis underscores that all variables, except 'Energy use' (EU), are considered significant predictors. The 'Energy use' variable, with a p -value of 0.42, falls outside the conventional significance threshold of 0.05, suggesting it does not significantly contribute to the model. Further evaluation of the model's overall efficacy is presented in Table 6.

Despite the model's high R-square value and a significant F-statistic, the large sums of squares suggest potential misspecification. To address this, we applied the Breusch–Pagan test to evaluate the presence of heteroskedasticity, which could indicate a preference for a random effects model over the pooled OLS model. This test is used to test for heteroskedasticity in the residuals. When applied to panel data, it tests the null hypothesis that variances across entities (countries) are equal. If the test is significant, it suggests the

presence of random effects, implying a random effects model might be more appropriate than a pooled OLS model. The hypotheses for the Breusch–Pagan test in this context are as follows:

- Null Hypothesis (H_0): There is no individual heteroskedasticity. This means that the variances across entities are equal, which implies that there are no individual-specific effects or that these effects are not correlated with other regressors.
- Alternative Hypothesis (H_1): There is individual heteroskedasticity. This implies the presence of individual-specific effects that might be correlated with other regressors.

The Breusch–Pagan test yielded a chi-square value of 3653.4 with a p -value $< 2.2 \times 10^{-16}$, compelling us to reject H_0 . This outcome points to heteroskedasticity and the potential suitability of a random effects model. Consequently, our next step involves a thorough investigation of random effects models to better cater to the data's characteristics.

Table 5. Pooled OLS Model Results.

Term	Estimate	Std. Error	t-Value	p-Value	Significance
(Intercept)	−22,963.0	14,169.0	−1.6207	0.1057332	
EPC	3.1854	0.82793	3.8475	0.0001351	***
EPF	−337.13	113.06	−2.9818	0.0030078	**
EU	−1.6263	2.0231	−0.8038	0.4218736	
F	−0.027955	0.00259	−10.7934	$< 2.2 \times 10^{-16}$	***
Fa	758.86	96.324	7.8782	2.153×10^{-14}	***
G	1991.4	591.33	3.3677	0.0008176	***
RE	−1466.4	309.84	−4.7327	0.000002902	***
TG	0.85062	0.0016949	501.8764	$< 2.2 \times 10^{-16}$	***

Significance codes: *** 0.001, ** 0.01.

Table 6. Pooled OLS Performance Summary.

Statistic	Value
Total Sum of Squares	6.78×10^{14}
Residual Sum of Squares	6.82×10^{11}
R-Squared	0.999
Adj. R-Squared	0.99898
F-statistic	61,022.1
p-value	$< 2.22 \times 10^{-16}$

5.1.2. Random Effects Models

In our exploration, we consider two distinct variants of random effects models. The initial model accommodates for unobserved heterogeneity among entities by treating these variations as random variables that are uncorrelated with the predictor variables. The subsequent model advances this approach by integrating a time trend, enabling the capture of systematic temporal changes impacting all entities. This modification aims to offer a more refined comprehension of the interplay between temporal dynamics and entity-specific variations. The formulas for these models are presented in the following. For the first model (random effects model without time trend), we have the following:

$$C_{oit} = \beta_0 + \beta_1 RE_{it} + \beta_2 Fa_{it} + \beta_3 EPF_{it} + \beta_4 G_{it} + \beta_5 TG_{it} + \beta_6 F_{it} + \beta_7 EU_{it} + \beta_8 EPC_{it} + u_i + \epsilon_{it} \quad (10)$$

For the second model (random effects model with time trend), we again have the following. In these equations, u_i denotes the unobserved random effect for entity i , γ_t represents the fixed effect for time t , and ϵ_{it} is the idiosyncratic error term. The comparison of these models is summarized in Table 7, highlighting their fit to the data.

$$C_{oit} = \beta_0 + \beta_1 RE_{it} + \beta_2 Fa_{it} + \beta_3 EPF_{it} + \beta_4 G_{it} + \beta_5 TG_{it} + \beta_6 F_{it} + \beta_7 EU_{it} + \beta_8 EPC_{it} + \gamma_t + u_i + \epsilon_{it} \quad (11)$$

Table 7. Comparison of random effects models.

Criteria	Model 1	Model 2
R-Squared	0.98216	0.98305
Adj. R-Squared	0.98187	0.98189
Total Sum of Squares	5.37×10^{12}	5.36×10^{12}
Residual Sum of Squares	9.59×10^{10}	9.09×10^{10}
AIC	10,972.74	10,994.36
BIC	11,010.67	11,133.44
Chisq	27,031.7	27,087.8
DF	8	32
<i>p</i> -Value	$<2.22 \times 10^{-16}$	$<2.22 \times 10^{-16}$

Although both models exhibit commendable data fit, the second model, with its marginally superior R-squared and lower residual sum of squares, suggests a nuanced understanding through the incorporation of a time trend. However, the increase in AIC and BIC values signals a caution against overcomplicating the model without clear justification. To address this, we have employed the Wald test to evaluate the following hypotheses:

- Null Hypothesis (H_0): The coefficients of time-specific effects are jointly zero, indicating the extended model does not significantly outperform the baseline model.
- Alternative Hypothesis (H_1): At least one time-specific effect coefficient is nonzero, suggesting the second model provides a superior fit.

Given an F-test value of 1.0544 and a *p*-value of 0.394, we fail to reject H_0 , indicating no substantial evidence that time-specific effects are significantly contributing to the model. This outcome favors the simpler random effects model without the time trend for its parsimony and effective representation of the data. The results of this simpler random effects model are presented in Table 8.

Table 8 highlights key findings regarding the significance of various characteristics in the context of CO₂ emissions. Notably, GDP, total greenhouse gas emission, and forest area emerge as significant predictors, each demonstrating a substantial role at the 0.05 level, with *p*-values of 0.005, 2.2×10^{-16} , and 1.987×10^{-14} , respectively. Other variables, showing *p*-values ranging from 0.11 to 0.43, suggest their non-significant impact at the conventional 5% significance level. This analysis prompts a focus on fixed effect models to examine entity-specific variations on CO₂ emissions, contrasting fixed characteristics against random, unobserved factors.

Table 8. Random effects model results.

Term	Estimate	Std. Error	z-Value	<i>p</i> -Value	Significance
Intercept	−62,514.000	22,289.000	−2.8046	0.005037	**
RE	462.66000	302.8000	1.5279	0.126534	
Fa	658.13000	458.42000	1.4356	0.151103	
EPF	247.48000	157.84000	1.5679	0.116907	
G	842.2000	305.46000	2.7571	0.005831	**
TG	0.920190	0.006387	144.0800	$<2.2 \times 10^{-16}$	***
F	−0.092565	0.012098	−7.6515	1.987×10^{-14}	***
EU	1.807400	2.331900	0.7751	0.438292	
EPC	−0.717160	0.677250	−1.0589	0.289628	

Significance codes: *** 0.001, ** 0.01.

5.1.3. Fixed Effects Models

This subsection explores the application of two distinct fixed effects models to our dataset. The first model implements a linear regression framework augmented by entity-specific fixed effects. This adjustment allows for the accommodation of unobserved heterogeneity among entities by introducing unique intercepts for each entity. The second model

employs a GLM framework, assuming a Gaussian distribution. This approach mirrors the structure of the first model but is designed to offer enhanced robustness in the presence of heteroscedasticity or non-normal error distributions. The equations governing these models are as follows:

$$C_{oit} = \alpha_i + \beta_1 RE_{it} + \beta_2 Fa_{it} + \beta_3 EPF_{it} + \beta_4 G_{it} + \beta_5 TG_{it} + \beta_6 F_{it} + \beta_7 EU_{it} + \beta_8 EPC_{it} + \epsilon_{it} \quad (12)$$

In the above equation, α_i symbolizes the entity-specific fixed effect for entity i , β values denote the coefficients of the respective predictors, and ϵ_{it} is the error term. Models D and E, while sharing the same structural formula, diverge in their estimation techniques and foundational assumptions. Model D employs dummy variables for entity-specific fixed effects, offering precise capture of entity variations. Model E, within a GLM framework and assuming a Gaussian distribution, seeks robustness under specific conditions. The choice between fixed and random effects models is guided by a Hausman test, with the following hypotheses:

- H_0 : The random effects estimator is consistent and efficient, implying no correlation between individual-specific effects and predictors.
- H_1 : The random effects estimator is inconsistent, favoring the fixed effects estimator due to correlations of individual-specific effects with predictors.

Rejecting H_0 (with p -value < 0.05) favors fixed effects models for our analysis. Comparisons based on AIC and BIC indicate Model E as the superior fit, underscored by its lower BIC and log-likelihood, indicating a close alignment between model predictions and actual data.

The examination of variable significance in Table 9, using a 0.05 significance level, identifies Renewable energy consumption, Electricity production from oil, gas, and coal sources, and GDP per unit of energy use as significantly positively impacting the dependent variable, with p -values of 0.01103, 0.01531, and 0.01301, respectively. Total greenhouse gas emission emerges as the most influential variable, showing a strong positive correlation with a p -value of less than 0.0001. Conversely, Forest Area (sq. km) significantly negatively impacts the dependent variable, also with a p -value of less than 0.0001. Variables such as Forest area (% of land area), Energy use, and Electric power consumption did not achieve statistical significance at the 0.05 level. Given the high correlation coefficient (-0.83 in Figure 5) between Renewable energy consumption and Electricity production from oil, gas, and coal sources, and the lower p -value of Renewable energy consumption, only the latter was included in subsequent analyses to address multicollinearity concerns. This approach underscores the importance of considering both model fit and variable significance, guiding a robust understanding of the determinants of CO₂ emissions.

Table 9. Fixed effects model results.

Variable	Estimate	Standard Error	t-Value	p-Value	Significance
Renewable energy consumption	7.40×10^2	2.90×10^2	2.5518	0.011	*
Forest area (% of land area)	-2.64×10^2	1.50×10^3	-0.1761	0.8603	
Electricity production from oil, gas, and coal sources	3.73×10^2	1.53×10^2	2.4337	0.015	*
GDP per unit of energy use	8.74×10^2	3.51×10^2	2.4928	0.013	*
Total greenhouse gas emission	9.63×10^{-1}	7.54×10^{-3}	127.77	<0.0001	***
Forest area (sq. km)	-5.46×10^{-1}	9.73×10^{-2}	-5.616	<0.0001	***
Energy use	-1.95×10^0	2.32×10^0	-0.843	0.3994	
Electric power consumption	3.60×10^{-2}	6.56×10^{-1}	0.055	0.9563	

Significance codes: *** 0.001, * 0.05.

5.2. Phase II: Prediction and Cluster Analysis of CO₂ Emissions

This phase analyzes and predicts CO₂ emissions across nations by utilizing the features identified in phase I. This section focuses on predicting emission over a three-year period. Following these predictions, countries are clustered by their CO₂ emissions trends, and the clusters have been then analyzed to gain a deeper understanding of different emission patterns using the most significant indicators identified in phase I.

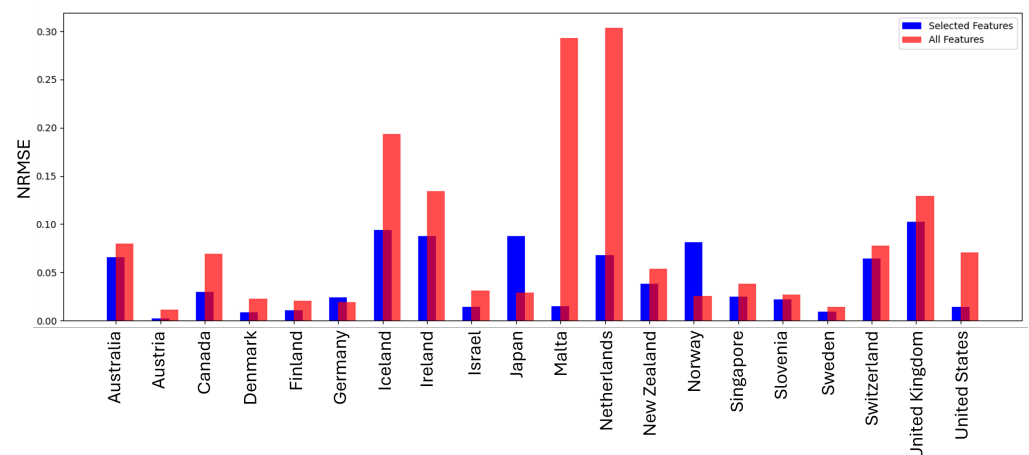
5.2.1. Prediction of CO₂ Emissions

Based upon the key characteristics identified in phase I, such as renewable energy consumption, GDP per unit of energy use, greenhouse gas emission, and forest area, CO₂ emissions are predicted from 2012 to 2014. It is important to note that while the dataset provided by the World Development Indicators (WDI) database offers comprehensive data for the period from 1990 to 2023, key indicators for more recent years, particularly post-2015, were not available for most high-HDI countries. To ensure the robustness of our predictions and avoid inaccuracies due to incomplete data, we trained the model on the available data from 1990 to 2011, and the forecast was made for the period 2012 to 2014. The prediction of CO₂ emissions is performed using the SARIMAX model where prediction results from two different scenarios are presented. In the first scenario, we predict CO₂ emissions over a three-year period using all the eight features of CO₂ indicators. In the second scenario, the prediction with SARIMAX over the same three-year period is performed using the four selected features obtained from phase I implementing the best model (Model E). Table 10 highlights the RMSE and MAE of the SARIMAX model for both these scenarios, whereas Figure 9 illustrates the NRMSE across different countries, comparing the outcomes using all the available features with the features selected from phase I.

The results presented in Table 10 indicate that the SARIMAX model benefits significantly from utilizing selected features from phase I in predicting CO₂ emissions. Switzerland, in particular, experienced a 31.5% decrease in RMSE and a 36.0% decrease in MAE, demonstrating the effectiveness of the feature selection. For Australia, RMSE and MAE decreased to 8693.57 and 8692.87, respectively, from 24,051.84 and 23,682.12, highlighting significant improvements. These results suggest that better emission estimates can be achieved using a more focused set of features. As shown in Figure 9, the NRMSE across various countries exhibits notable reductions when the selected features are applied. For instance, Norway's NRMSE decreased by approximately 50%, indicating a marked improvement in model accuracy. Although the selected features did not yield uniformly lower NRMSE values for all countries, such as Germany and Japan, the overall trend supports the hypothesis that strategic feature selection significantly enhances emission prediction. By examining three representative countries categorized according to their relative HDI positions, we gain a deeper understanding of the model's predictive performance. Figure 10 compares the model's performance using the entire feature set against predictions made with the selected phase I features. This comparison shows that using selected features generally enhances the model's performance, as demonstrated by closer approximations to actual CO₂ emissions for these countries. An in-depth examination of the Netherlands' predictive results highlights the approach's effectiveness. SARIMAX, using its entire feature set, incorrectly projects an increase in CO₂ emissions from 2011 to 2012, contradicting the actual emission decline. In contrast, when the selected phase I features are applied, the model's prediction aligns more closely with the actual downward trend, emphasizing the importance of feature selection in improving prediction accuracy.

Table 10. RMSE and MAE of SARIMAX model predictions using all features versus selected features from phase I.

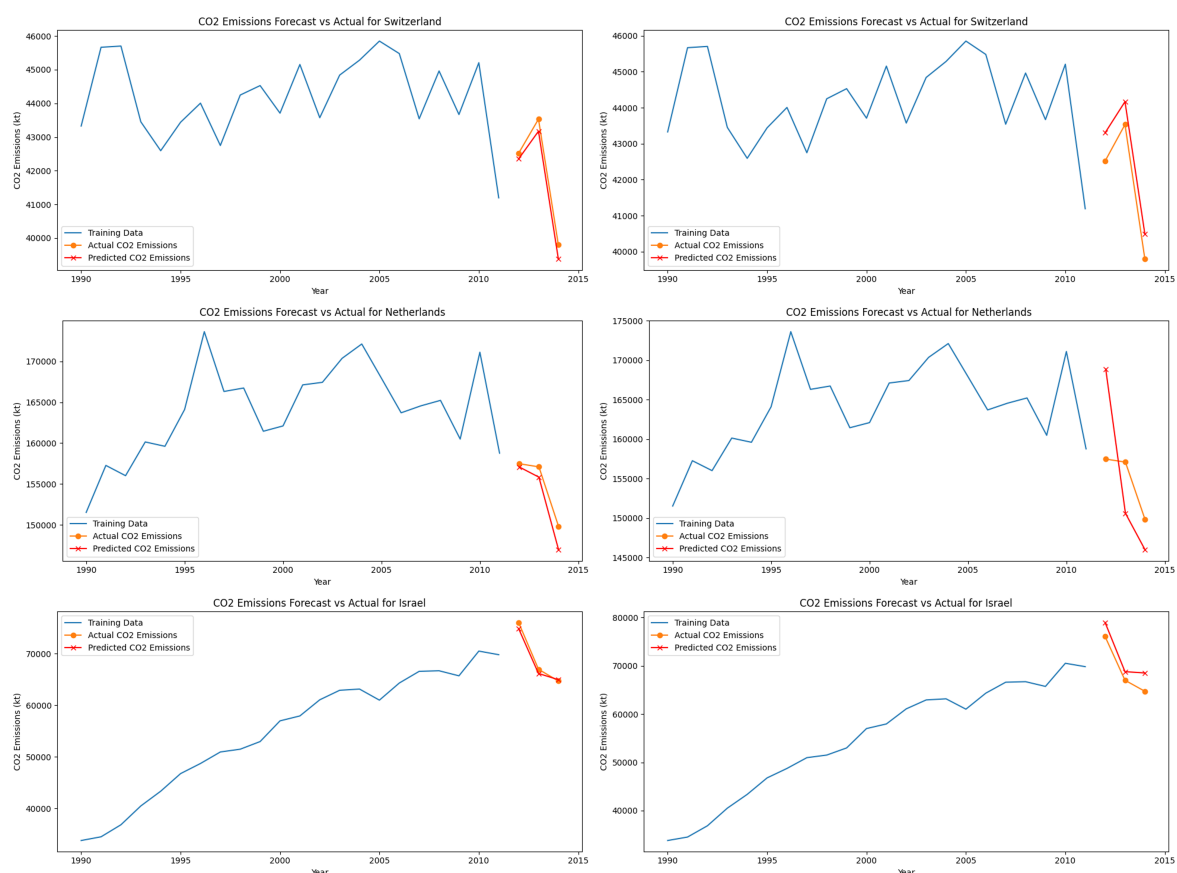
Country	All Features		Selected Features	
	RMSE	MAE	RMSE	MAE
Switzerland	435.46	433.32	298.27	277.25
Norway	321.19	318.40	1036.16	903.42
Iceland	95.76	95.59	46.47	37.64
Australia	24,051.84	23,682.12	8693.57	8692.87
Denmark	699.83	541.63	250.14	242.44
Sweden	328.78	267.76	189.28	183.01
Ireland	1641.91	1444.42	1428.17	1401.92
Germany	4612.22	4328.02	5268.25	4508.37
Netherlands	6716.53	6069.93	1501.10	1382.38
Finland	398.62	374.73	199.88	190.58
Singapore	648.48	586.43	389.72	334.27
New Zealand	793.51	676.55	456.41	435.40
Canada	10,962.22	10,893.59	4654.82	3003.13
United Kingdom	16,159.76	13,459.35	12,826.57	9421.95
Japan	9982.56	7788.30	12,081.59	11,135.11
Israel	1284.12	1122.49	513.74	424.41
United States	68,695.06	61,815.72	13,433.16	12,225.59
Malta	206.78	157.41	10.28	7.77
Slovenia	170.08	144.07	101.13	90.83
Austria	209.25	197.29	41.04	36.72

**Figure 9.** NRMSE comparison across countries using all features versus selected features from phase I.

To illustrate the efficacy of SARIMAX, we compared its performance with a recurrent neural network (RNN) and its variant, the long short-term memory (LSTM) network. As shown in Table 11, SARIMAX outperformed both RNN and LSTM in terms of NRMSE for the 3-year forecasting period across multiple countries with all features. SARIMAX's ability to explicitly model seasonality and trends in CO₂ emissions data, along with incorporating external regressors, allows it to more accurately capture short-term emission patterns. In contrast, RNN and LSTM, while adept at handling complex temporal dependencies, require longer time horizons and larger datasets to realize their full potential. For this study, SARIMAX provided superior short-term forecasting accuracy, making it the most appropriate model for the analysis. The decision to use a 3-year forecasting horizon was based on its balance between capturing short-term trends and providing sufficient temporal variation for meaningful predictions. Shorter periods, such as 1 year, may not capture emission dynamics, while longer periods, like 5 years, introduce greater uncertainty and variability, especially for countries with complex, non-linear emission patterns. Additionally, limited data for longer periods reduces forecast reliability.

Table 11. Comparison of NRMSE for SARIMAX, RNN, and LSTM models across countries.

Country	NRMSE		
	RNN	LSTM	SARIMAX
Switzerland	0.442783	0.480825	0.077552
Norway	0.121911	0.085626	0.025110
Iceland	0.344458	0.475491	0.193595
Australia	0.101343	0.160552	0.079787
Denmark	0.247964	0.265102	0.022905
Sweden	0.161550	0.245968	0.013986
Ireland	0.556493	0.533763	0.133853
Germany	0.133541	0.121431	0.019064
Netherlands	0.444045	0.433081	0.303878
Finland	0.492177	0.422556	0.020518
Singapore	0.253100	0.271664	0.037952
New Zealand	0.268500	0.181010	0.053941
Canada	0.118399	0.122721	0.069235
United Kingdom	0.424373	0.336663	0.129244
Japan	0.404918	0.379645	0.028718
Israel	0.122919	0.153771	0.031110
United States	0.335382	0.490067	0.070943
Malta	0.238182	0.342186	0.293078
Slovenia	0.424588	0.513795	0.026756
Austria	0.376747	0.321901	0.011265

**Figure 10.** SARIMAX model predictions for three representative countries using selected features from phase I (left) and all features (right).

5.2.2. Cluster Analysis

After predicting CO₂ emissions, in this subsection, we perform a cluster analysis by applying the DTW algorithm, which categorizes the countries based on their CO₂ emission trends. This methodology, which builds on insights obtained in the data understanding

phase, effectively groups countries with similar emission patterns into three clusters, as shown in Figure 11.

In Figure 11, along the x-axis, we have time spanning from the year 1990 to 2014, and the y-axis indicates the standardized CO₂ emissions levels for each country. By standardizing the emission trends, it is possible to observe patterns and dynamics of CO₂ emissions for each country more easily. The first cluster, as shown in Figure 11a includes countries such as Switzerland and the United Kingdom, whose emission patterns exhibit notable fluctuations but tend to converge in later years towards the cluster center. In addition to having a variety of industrial activities, this group may also reflect economies that have progressively adopted emission reduction strategies. As illustrated in the Figure 11b, Norway, Australia, and Canada are included in the second cluster. These nations share a consistent upward trend in CO₂ emissions before stabilizing or slightly decreasing as they approach later times. This pattern may indicate sustained growth in emission caused by economic expansion, followed by a shift toward stabilization, possibly influenced by environmental policies. A third and final cluster can be seen in Figure 11c, which includes the United States and the Netherlands, among other countries. A trend of increasing emissions is observed in this diverse group until 2004, following which a gradual decline is observed. These trends may be associated with a peak in industrial emission followed by concerted efforts to reduce emission and adopt renewable energy sources. During the period examined, there were distinct trajectories in the emission of CO₂ in these countries, which may reflect the different socio-economic and environmental policies of each country. A comprehensive evaluation of the robustness of our clustering approach is provided by Figures 12 and 13, which provide insight into how CO₂ emissions trends are aligned. In Figure 12, the distances between cluster centers are presented, revealing inter-cluster variability, whereas in Figure 13, an individual country trend is compared to the central trend of the cluster, revealing intra-cluster consistency.

DTW visualizations provide a robust analytical tool for validating CO₂ emissions trends clustering across various countries. Figure 12a–c demonstrate the temporal alignment between cluster centers, illustrating disparities between clusters. In each heatmap, the x-axis shows the sequential time step corresponding to the standardized emission values for one cluster, while the y-axis relates to the time step corresponding to another cluster. In these figures, a gradient of colors indicates local point-to-point distances, with warmer colors indicating higher distance and cooler colors showing closer proximity, which illustrates the degree of similarity between the two time series. There is notable variation in the optimal warping paths represented by the red lines, with deviations from the diagonal indicating periods of divergence in CO₂ emissions trends between the clusters.

Additionally, Figure 13a–c examine the alignment between emission trends within individual countries and cluster centers, thus examining the coherence of countries within their clusters. Each heatmap's x-axis represents the central trend of the cluster, while its y-axis corresponds to the central trend of a particular country within that cluster. A strong alignment with the cluster's central trend is indicated by the proximity of the red lines to the diagonal, which indicates that the optimal warping paths are aligned. Overall DTW distances at the top of each heatmap provide a quantitative measure of the alignment, and the small distances indicate that the clustering is reliable. In Table 12, different clusters are analyzed based on the selected features from phase I.

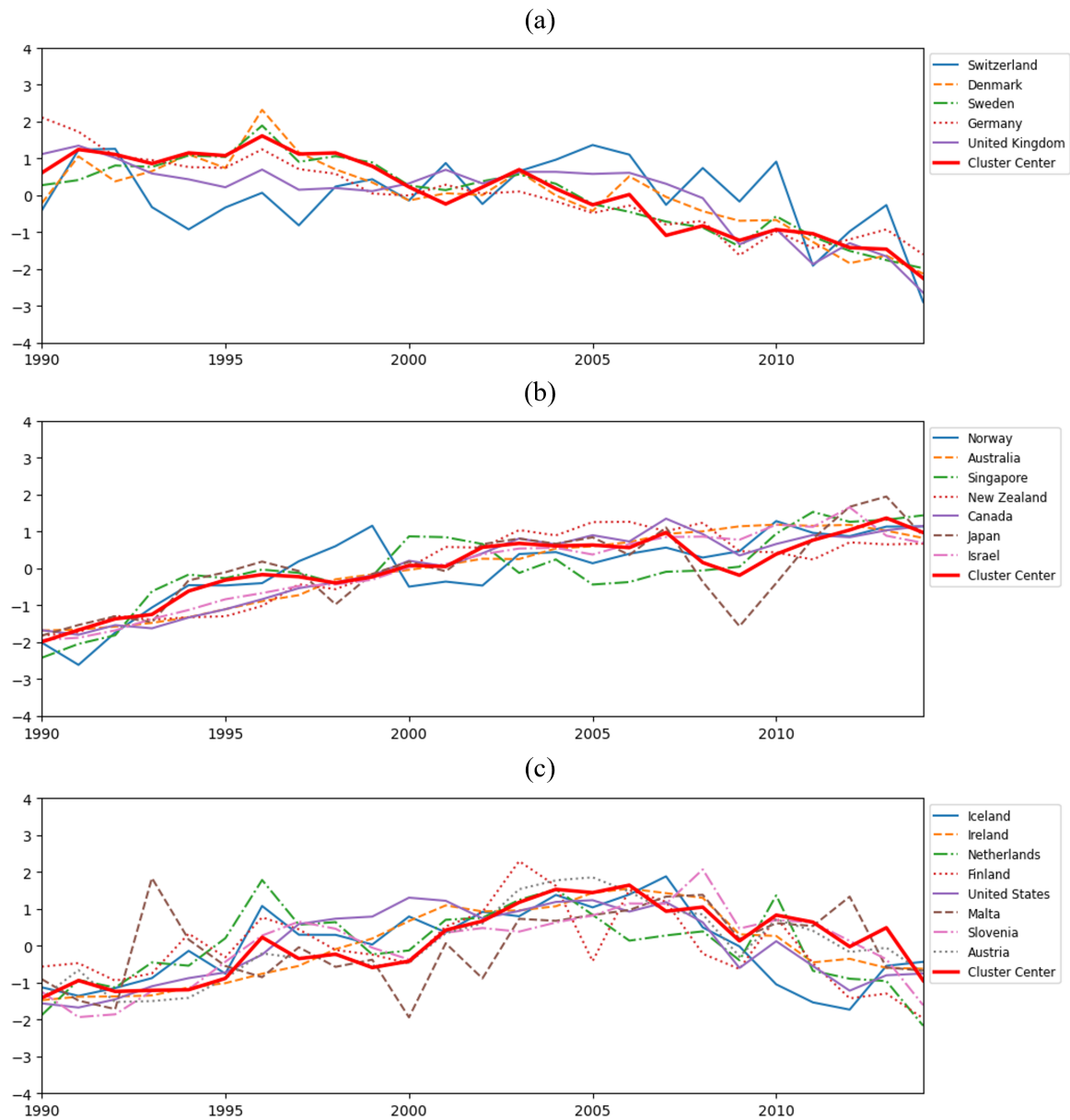


Figure 11. Cluster analysis of the countries based on carbon emissions patterns ((a): converging fluctuations, (b): upward trends stabilizing, (c): peak followed by decline).

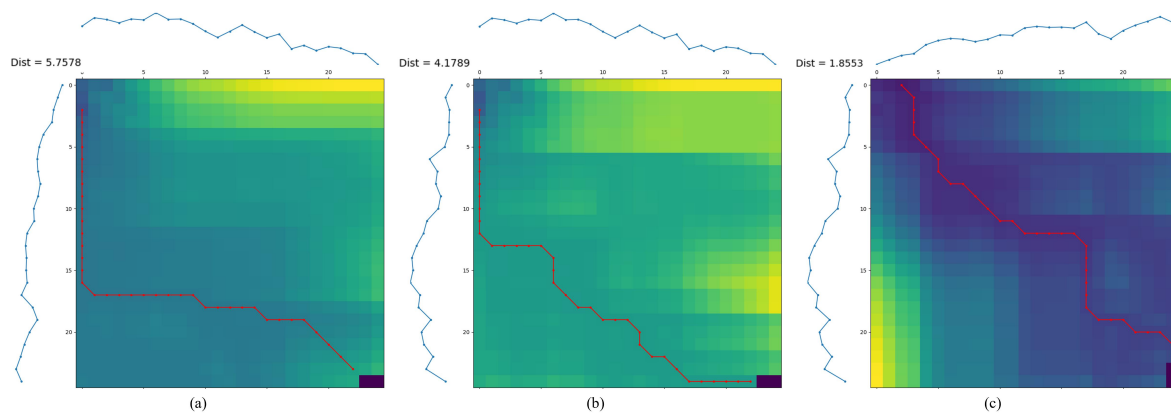


Figure 12. Dynamic time warping distances between cluster centers for CO₂ emissions.

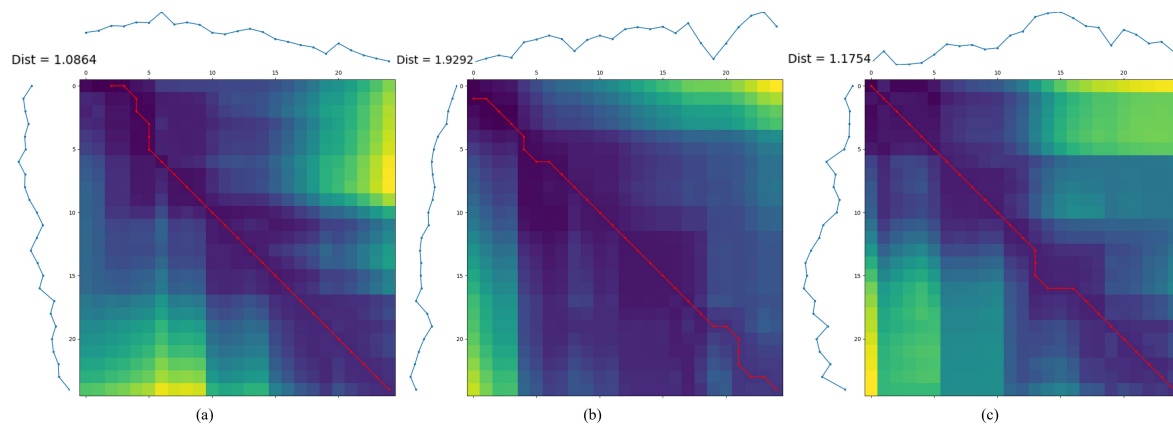


Figure 13. Alignment of individual country CO₂ emissions trends with cluster central trends.

Table 12. Analysis of the clusters based on the selected features.

Feature	Cluster	Mean	Median	Standard Deviation	Annual Growth Rate (%)
Renewable energy consumption	Cluster 1	15.95	13.64	13.57	−3.86
	Cluster 2	18.1	8.42	18.91	−10.91
	Cluster 3	18.78	9.73	20.52	−1.76
Electricity production from oil, gas and coal sources	Cluster 1	43.37	61.03	34.41	17.12
	Cluster 2	57.33	61.18	37.66	35.07
	Cluster 3	54.67	53.39	33.2	22.61
GDP per unit of energy use	Cluster 1	8.65	8.12	3.6	2.7
	Cluster 2	6.95	6.34	2.97	5.21
	Cluster 3	6.88	6.23	3.53	6.93
Total greenhouse gas emission	Cluster 1	364,635.94	73,846.74	385,349.73	9.94
	Cluster 2	389,649.8	83,368.13	432,138.17	2.59
	Cluster 3	855,195.2	69,949.81	2,102,668.71	13.84
Forest area	Cluster 1	88,400.66	29,750.0	104,200.41	4.26
	Cluster 2	751,927.56	121,149.0	1,195,835.76	−16.44
	Cluster 3	417,152.51	9722.06	1,001,584.13	25.37

As indicated by the moderate and declining growth rate of renewable energy consumption in cluster 1, which can be referred to as the “Reducing emission Group”, CO₂ emissions were decreasing. Despite the significant increase in the annual growth rate of electricity produced by oil, gas, and coal, the mean electricity production from these sources is 43.37 percent, suggesting a transition towards cleaner energy sources has occurred. A high GDP per unit of energy use, as well as a growth in forest area, support the hypothesis that the countries belonging to this cluster have moved towards sustainable practices. As a result of an increase in CO₂ emissions, cluster 2 can be identified as a group with “Growing emission”. This is consistent with the high mean of 57.33% in electricity production from oil, gas, and coal sources and the highest growth rate of 35.07%, which indicates a heavy reliance on non-renewable energy sources. Although the cluster is characterized by the largest forest area, its negative annual growth rate is concerning, potentially exacerbating its overall environmental impact. In cluster 3, the “Peak and Decline emission Group”, CO₂ emissions initially increased, followed by a decline, resulting in the largest mean renewable energy consumption at 18.78% and a less pronounced decrease in growth rate. As indicated by the highest annual growth rate in forest area, which is 25.37 percent, significant total greenhouse gas emission are aligned with this pattern but are attempting to reverse the trend. A summary of the results over the period from 1990 to 2014 suggests that cluster 1 made significant progress toward reducing emission. Cluster 2 experienced an upward trend in emission, indicating a growing environmental challenge during this

period. Meanwhile, cluster 3 appeared to correct its initially high emissions, demonstrating a shift towards more sustainable practices in the latter part of the timeframe. These historical trends, analyzed in conjunction with the detailed characteristics identified in phase I, provide a comprehensive perspective on the environmental trajectories of the clusters and the effectiveness of their energy and environmental policies.

6. Conclusions

In the context of addressing global climate change, this research marks a significant step forward by pinpointing the factors that drive CO₂ emissions. By combining a broad range of variables with diverse statistical models, this study provides an in-depth and comprehensive analysis. Its standout feature is a two-phase methodology that focuses on analyzing key features and enhancing prediction precision, leading to notably improved forecasting accuracy. This approach ensures predictions mirror actual trends in CO₂ emissions, whether they are rising or falling, thus making future projections more dependable. A key methodological breakthrough of this research is the application of the DTW algorithm during its second phase, enabling the categorization of countries into clusters based on their emission trajectories. This provides deep insights into the global landscape of CO₂ emissions, highlighting both the commonalities and disparities across nations.

This study has substantial practical implications, especially in the development of policies. By providing precise forecasts of future emission trends and a detailed understanding of emission dynamics on a country-by-country basis, it empowers policymakers and environmental strategists to devise climate policies that are both impactful and finely tuned to specific needs. Thus, the study's contributions extend far beyond academic circles, offering practical tools for combating climate change. It aids in informed policymaking and strategic environmental planning, significantly aiding the global initiative to curb CO₂ emissions and counteract climate change's negative impacts. Although the models aligned closely with actual emission trends, the analysis was limited to available historical data. Future studies could explore external validation, such as cross-validation or out-of-sample testing with more available data.

Author Contributions: Conceptualization: I.A., H.K. and F.I.; methodology: H.K., F.I. and I.A.; formal analysis and investigation: H.K., F.I. and A.S.R.; writing—original draft preparation: H.K., F.I. and A.S.R.; writing—review and editing: A.N. and I.A.; supervision: A.N. and I.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been generously supported by the United States Environmental Protection Agency (EPA) under the Pollution Prevention (P2) practices grant.

Data Availability Statement: The original data presented in the study are openly available at <https://databank.worldbank.org/source/world-development-indicators> (accessed on 15 January 2024).

Acknowledgments: The authors would like to thank the Environmental Protection Agency (EPA) and West Virginia University (WVU) for their support.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Hao, Y.; Chen, H.; Wei, Y.M.; Li, Y.M. The influence of climate change on CO₂ (carbon dioxide) emissions: An empirical estimation based on Chinese provincial panel data. *J. Clean. Prod.* **2016**, *131*, 667–677. [CrossRef]
2. Pettoirelli, N.; Graham, N.A.; Seddon, N.; Maria da Cunha Bustamante, M.; Lowton, M.J.; Sutherland, W.J.; Koldewey, H.J.; Prentice, H.C.; Barlow, J. Time to integrate global climate change and biodiversity science-policy agendas. *J. Appl. Ecol.* **2021**, *58*, 2384–2393. [CrossRef]

3. Lynn, J.; Peeva, N. Communications in the IPCC's Sixth Assessment Report cycle. *Clim. Chang.* **2021**, *169*, 18. [[CrossRef](#)] [[PubMed](#)]
4. Liu, D.; Guo, X.; Xiao, B. What causes growth of global greenhouse gas emissions? Evidence from 40 countries. *Sci. Total Environ.* **2019**, *661*, 750–766. [[CrossRef](#)]
5. Jones, M.W.; Peters, G.P.; Gasser, T.; Andrew, R.M.; Schwingshackl, C.; Gütschow, J.; Houghton, R.A.; Friedlingstein, P.; Pongratz, J.; Le Quéré, C. National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850. *Sci. Data* **2023**, *10*, 155. [[CrossRef](#)]
6. Zheng, X.; Streimikiene, D.; Balezentis, T.; Mardani, A.; Cavallaro, F.; Liao, H. A review of greenhouse gas emission profiles, dynamics, and climate change mitigation efforts across the key climate change players. *J. Clean. Prod.* **2019**, *234*, 1113–1133. [[CrossRef](#)]
7. Gür, T.M. Carbon dioxide emissions, capture, storage and utilization: Review of materials, processes and technologies. *Prog. Energy Combust. Sci.* **2022**, *89*, 100965. [[CrossRef](#)]
8. Yoro, K.O.; Daramola, M.O. CO₂ emission sources, greenhouse gases, and the global warming effect. In *Advances in Carbon Capture*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 3–28.
9. Loucks, D.P. Impacts of climate change on economies, ecosystems, energy, environments, and human equity: A systems perspective. In *The Impacts of Climate Change*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 19–50.
10. Fekete, H.; Kuramochi, T.; Roelfsema, M.; den Elzen, M.; Forsell, N.; Höhne, N.; Luna, L.; Hans, F.; Sterl, S.; Olivier, J.; et al. A review of successful climate change mitigation policies in major emitting economies and the potential of global replication. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110602. [[CrossRef](#)]
11. Seo, S.N. Beyond the Paris Agreement: Climate change policy negotiations and future directions. *Reg. Sci. Policy Pract.* **2017**, *9*, 121–140. [[CrossRef](#)]
12. Chen, C.C. Comparative impacts of energy sources on environmental quality: A five-decade analysis of Germany's Energiewende. *Energy Rep.* **2024**, *11*, 3550–3561. [[CrossRef](#)]
13. Almström, P.; Anderstig, C.; Sundberg, M. Effects on sectors and regions of a carbon tax increase in Sweden: Analysis with an SCGE model. *Ann. Reg. Sci.* **2024**, *73*, 731–756. [[CrossRef](#)]
14. Andoh, J.; Oduro, K.A.; Park, J.; Lee, Y. Towards REDD+ implementation: Deforestation and forest degradation drivers, REDD+ financing, and readiness activities in participant countries. *Front. For. Glob. Chang.* **2022**, *5*, 957550. [[CrossRef](#)]
15. Rakatama, A.; Pandit, R.; Ma, C.; Iftekhhar, S. The costs and benefits of REDD+: A review of the literature. *For. Policy Econ.* **2017**, *75*, 103–111. [[CrossRef](#)]
16. Cai, Y.; Sam, C.Y.; Chang, T. Nexus between clean energy consumption, economic growth and CO₂ emissions. *J. Clean. Prod.* **2018**, *182*, 1001–1011. [[CrossRef](#)]
17. Li, Z.; Wu, H.; Wu, F. Impacts of urban forms and socioeconomic factors on CO₂ emissions: A spatial econometric analysis. *J. Clean. Prod.* **2022**, *372*, 133722. [[CrossRef](#)]
18. Puertas, R.; Marti, L. Eco-innovation and determinants of GHG emissions in OECD countries. *J. Clean. Prod.* **2021**, *319*, 128739. [[CrossRef](#)]
19. Jiang, X.T.; Wang, Q.; Li, R. Investigating factors affecting carbon emission in China and the USA: A perspective of stratified heterogeneity. *J. Clean. Prod.* **2018**, *199*, 85–92. [[CrossRef](#)]
20. Wang, P.; Wu, P.; Song, Y.; Hampson, K.; Zhong, Y. A novel spatio-temporally stratified heterogeneity model for identifying factors influencing carbon emissions. *Energy Build.* **2023**, *280*, 112714. [[CrossRef](#)]
21. Adekoya, O.B.; Olabode, J.K.; Rafi, S.K. Renewable energy consumption, carbon emissions and human development: Empirical comparison of the trajectories of world regions. *Renew. Energy* **2021**, *179*, 1836–1848. [[CrossRef](#)]
22. Hossain, M.A.; Chen, S. Nexus between Human Development Index (HDI) and CO₂ emissions in a developing country: Decoupling study evidence from Bangladesh. *Environ. Sci. Pollut. Res.* **2021**, *28*, 58742–58754. [[CrossRef](#)]
23. Mohammed, A.; Li, Z.; Arowolo, A.O.; Su, H.; Deng, X.; Najmuddin, O.; Zhang, Y. Driving factors of CO₂ emissions and nexus with economic growth, development and human health in the Top Ten emitting countries. *Resour. Conserv. Recycl.* **2019**, *148*, 157–169. [[CrossRef](#)]
24. Ahmad, T.; Zhang, D. A critical review of comparative global historical energy consumption and future demand: The story told so far. *Energy Rep.* **2020**, *6*, 1973–1991. [[CrossRef](#)]
25. Spierre, S.G.; Seager, T.P.; Selinger, E. The diminishing returns to the 2010 human Development Index: Implications for climate policy and sustainable development. *J. Sustain. Dev.* **2013**, *6*, 34. [[CrossRef](#)]
26. Ambrósio, G.; Da Cunha, D.A.; Pires, M.V.; Costa, L.; Faria, R.M.; Gurgel, A.C. Human development, greenhouse gas emissions and sub-national mitigation burdens: A Brazilian perspective. *Discov. Sustain.* **2021**, *2*, 35. [[CrossRef](#)]
27. Wong, Y.F.; Ankervold, S. Comparative Analysis of the Impact of Pollution on Human Development Index: High vs. Low Pollution Countries. Master's Thesis, UIS, Stavanger, Norway, 2024.
28. Hashmat, A.; Ghouse, G.; Ahmad, N. Impact of Economic and Environmental Dynamics on Human Development: An Analysis of HDI Classifications. *Bull. Bus. Econ. (BBE)* **2023**, *12*, 738–751. [[CrossRef](#)]

29. Mardani, A.; Liao, H.; Nilashi, M.; Alrasheedi, M.; Cavallaro, F. A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. *J. Clean. Prod.* **2020**, *275*, 122942. [\[CrossRef\]](#)
30. Le Quéré, C.; Korsbakken, J.I.; Wilson, C.; Tosun, J.; Andrew, R.; Andres, R.J.; Canadell, J.G.; Jordan, A.; Peters, G.P.; van Vuuren, D.P. Drivers of declining CO₂ emissions in 18 developed economies. *Nat. Clim. Chang.* **2019**, *9*, 213–217. [\[CrossRef\]](#)
31. Hao, Y. Effect of economic indicators, renewable energy consumption and human development on climate change: An empirical analysis based on panel data of selected countries. *Front. Energy Res.* **2022**, *10*, 841497. [\[CrossRef\]](#)
32. Fan, W.; Hao, Y. An empirical research on the relationship amongst renewable energy consumption, economic growth and foreign direct investment in China. *Renew. Energy* **2020**, *146*, 598–609. [\[CrossRef\]](#)
33. Rahman, Z.U.; Khattak, S.I.; Ahmad, M.; Khan, A. A disaggregated-level analysis of the relationship among energy production, energy consumption and economic growth: Evidence from China. *Energy* **2020**, *194*, 116836. [\[CrossRef\]](#)
34. Xie, Q.; Wang, X.; Cong, X. How does foreign direct investment affect CO₂ emissions in emerging countries? New findings from a nonlinear panel analysis. *J. Clean. Prod.* **2020**, *249*, 119422. [\[CrossRef\]](#)
35. Zhang, J.; Zhang, Y. Examining the economic and environmental effects of emissions policies in China: A Bayesian DSGE model. *J. Clean. Prod.* **2020**, *266*, 122026. [\[CrossRef\]](#)
36. Ding, Y.; Li, F. Examining the effects of urbanization and industrialization on carbon dioxide emission: Evidence from China's provincial regions. *Energy* **2017**, *125*, 533–542. [\[CrossRef\]](#)
37. Li, L.; Hong, X.; Tang, D.; Na, M. GHG emissions, economic growth and urbanization: A spatial approach. *Sustainability* **2016**, *8*, 462. [\[CrossRef\]](#)
38. Sikder, M.; Wang, C.; Yao, X.; Huai, X.; Wu, L.; KwameYeboah, F.; Wood, J.; Zhao, Y.; Dou, X. The integrated impact of GDP growth, industrialization, energy use, and urbanization on CO₂ emissions in developing countries: Evidence from the panel ARDL approach. *Sci. Total Environ.* **2022**, *837*, 155795. [\[CrossRef\]](#)
39. Arneth, A.; Sitch, S.; Pongratz, J.; Stocker, B.D.; Ciais, P.; Poulter, B.; Bayer, A.D.; Bondeau, A.; Calle, L.; Chini, L.P.; et al. Historical carbon dioxide emissions caused by land-use changes are possibly larger than assumed. *Nat. Geosci.* **2017**, *10*, 79–84. [\[CrossRef\]](#)
40. Popp, A.; Krause, M.; Dietrich, J.P.; Lotze-Campen, H.; Leimbach, M.; Beringer, T.; Bauer, N. Additional CO₂ emissions from land use change—Forest conservation as a precondition for sustainable production of second generation bioenergy. *Ecol. Econ.* **2012**, *74*, 64–70. [\[CrossRef\]](#)
41. Raihan, A.; Begum, R.A.; Nizam, M.; Said, M.; Pereira, J.J. Dynamic impacts of energy use, agricultural land expansion, and deforestation on CO₂ emissions in Malaysia. *Environ. Ecol. Stat.* **2022**, *29*, 477–507. [\[CrossRef\]](#)
42. Bilgili, F.; Balsalobre-Lorente, D.; Kuşkaya, S.; Alnour, M.; Önderol, S.; Hoque, M.E. Are research and development on energy efficiency and energy sources effective in the level of CO₂ emissions? Fresh evidence from EU data. *Environ. Dev. Sustain.* **2024**, *26*, 24183–24219. [\[CrossRef\]](#)
43. Nejat, P.; Jomehzadeh, F.; Taheri, M.M.; Gohari, M.; Majid, M.Z.A. A global review of energy consumption, CO₂ emissions and policy in the residential sector (with an overview of the top ten CO₂ emitting countries). *Renew. Sustain. Energy Rev.* **2015**, *43*, 843–862. [\[CrossRef\]](#)
44. Shah, W.U.H.; Hao, G.; Yan, H.; Zhu, N.; Yasmeen, R.; Dincă, G. Role of renewable, non-renewable energy consumption and carbon emission in energy efficiency and productivity change: Evidence from G20 economies. *Geosci. Front.* **2023**, *15*, 101631. [\[CrossRef\]](#)
45. Hashmi, R.; Alam, K. Dynamic relationship among environmental regulation, innovation, CO₂ emissions, population, and economic growth in OECD countries: A panel investigation. *J. Clean. Prod.* **2019**, *231*, 1100–1109. [\[CrossRef\]](#)
46. Ahmadi, M.H.; Jashnani, H.; Chau, K.W.; Kumar, R.; Rosen, M.A. Carbon dioxide emissions prediction of five Middle Eastern countries using artificial neural networks. *Energy Sources Part A Recovery Util. Environ. Eff.* **2023**, *45*, 9513–9525. [\[CrossRef\]](#)
47. Fang, D.; Zhang, X.; Yu, Q.; Jin, T.C.; Tian, L. A novel method for carbon dioxide emission forecasting based on improved Gaussian processes regression. *J. Clean. Prod.* **2018**, *173*, 143–150. [\[CrossRef\]](#)
48. Li, X.; Zhao, Z.; Zhao, Y.; Zhou, S.; Zhang, Y. Prediction of energy-related carbon emission intensity in China, America, India, Russia, and Japan using a novel self-adaptive grey generalized Verhulst model. *J. Clean. Prod.* **2023**, *423*, 138656. [\[CrossRef\]](#)
49. Wang, Q.; Li, S.; Pisarenko, Z. Modeling carbon emission trajectory of China, US and India. *J. Clean. Prod.* **2020**, *258*, 120723. [\[CrossRef\]](#)
50. Adebayo, T.S.; Beton Kalmaz, D. Determinants of CO₂ emissions: Empirical evidence from Egypt. *Environ. Ecol. Stat.* **2021**, *28*, 239–262. [\[CrossRef\]](#)
51. Dong, K.; Dong, X.; Dong, C. Determinants of the global and regional CO₂ emissions: What causes what and where? *Appl. Econ.* **2019**, *51*, 5031–5044. [\[CrossRef\]](#)
52. Zhou, C.; Wang, S. Examining the determinants and the spatial nexus of city-level CO₂ emissions in China: A dynamic spatial panel analysis of China's cities. *J. Clean. Prod.* **2018**, *171*, 917–926. [\[CrossRef\]](#)
53. Jiang, J.; Ye, B.; Liu, J. Research on the peak of CO₂ emissions in the developing world: Current progress and future prospect. *Appl. Energy* **2019**, *235*, 186–203. [\[CrossRef\]](#)

54. Van Soest, H.L.; den Elzen, M.G.; van Vuuren, D.P. Net-zero emission targets for major emitting countries consistent with the Paris Agreement. *Nat. Commun.* **2021**, *12*, 2140. [[CrossRef](#)] [[PubMed](#)]
55. Olabi, A.; Obaideen, K.; Elsaid, K.; Wilberforce, T.; Sayed, E.T.; Maghrabie, H.M.; Abdelkareem, M.A. Assessment of the pre-combustion carbon capture contribution into sustainable development goals SDGs using novel indicators. *Renew. Sustain. Energy Rev.* **2022**, *153*, 111710. [[CrossRef](#)]
56. Bennedsen, M.; Hillebrand, E.; Koopman, S.J. Modeling, forecasting, and nowcasting US CO₂ emissions using many macroeconomic predictors. *Energy Econ.* **2021**, *96*, 105118. [[CrossRef](#)]
57. Wang, Z.; Jiang, Q.; Dong, K.; Mubarik, M.S.; Dong, X. Decomposition of the US CO₂ emissions and its mitigation potential: An aggregate and sectoral analysis. *Energy Policy* **2020**, *147*, 111925. [[CrossRef](#)]
58. Tian, S.; Xu, Y.; Wang, Q.; Zhang, Y.; Yuan, X.; Ma, Q.; Chen, L.; Ma, H.; Liu, J.; Liu, C. Research on peak prediction of urban differentiated carbon emissions—A case study of Shandong Province, China. *J. Clean. Prod.* **2022**, *374*, 134050. [[CrossRef](#)]
59. Wang, H.; Zhang, Z. Forecasting Chinese provincial carbon emissions using a novel grey prediction model considering spatial correlation. *Expert Syst. Appl.* **2022**, *209*, 118261. [[CrossRef](#)]
60. Huber, J.; Lohmann, K.; Schmidt, M.; Weinhardt, C. Carbon efficient smart charging using forecasts of marginal emission factors. *J. Clean. Prod.* **2021**, *284*, 124766. [[CrossRef](#)]
61. Adedoyin, F.F.; Zakari, A. Energy consumption, economic expansion, and CO₂ emission in the UK: The role of economic policy uncertainty. *Sci. Total Environ.* **2020**, *738*, 140014. [[CrossRef](#)]
62. Sarkodie, S.A.; Strezov, V. Assessment of contribution of Australia's energy production to CO₂ emissions and environmental degradation using statistical dynamic approach. *Sci. Total Environ.* **2018**, *639*, 888–899. [[CrossRef](#)]
63. Javanmard, M.E.; Tang, Y.; Wang, Z.; Tontiwachwuthikul, P. Forecast energy demand, CO₂ emissions and energy resource impacts for the transportation sector. *Appl. Energy* **2023**, *338*, 120830. [[CrossRef](#)]
64. Jiang, Y.; Long, Y.; Liu, Q.; Dowaki, K.; Ihara, T. Carbon emission quantification and decarbonization policy exploration for the household sector—Evidence from 51 Japanese cities. *Energy Policy* **2020**, *140*, 111438. [[CrossRef](#)]
65. Dong, C.; Dong, X.; Jiang, Q.; Dong, K.; Liu, G. What is the probability of achieving the carbon dioxide emission targets of the Paris Agreement? Evidence from the top ten emitters. *Sci. Total Environ.* **2018**, *622*, 1294–1303. [[CrossRef](#)] [[PubMed](#)]
66. Bakır, H.; Ağbulut, Ü.; Gürel, A.E.; Yıldız, G.; Güvenç, U.; Soudagar, M.E.M.; Hoang, A.T.; Deepanraj, B.; Saini, G.; Afzal, A. Forecasting of future greenhouse gas emission trajectory for India using energy and economic indexes with various metaheuristic algorithms. *J. Clean. Prod.* **2022**, *360*, 131946. [[CrossRef](#)]
67. Ghalandari, M.; Forootan Fard, H.; Komeili Birjandi, A.; Mahariq, I. Energy-related carbon dioxide emission forecasting of four European countries by employing data-driven methods. *J. Therm. Anal. Calorim.* **2021**, *144*, 1999–2008. [[CrossRef](#)]
68. Alotaibi, A.A.; Alajlan, N. Using quantile regression to analyze the relationship between socioeconomic indicators and carbon dioxide emissions in G20 countries. *Sustainability* **2021**, *13*, 7011. [[CrossRef](#)]
69. Mardani, A.; Streimikiene, D.; Nilashi, M.; Arias Aranda, D.; Loganathan, N.; Jusoh, A. Energy consumption, economic growth, and CO₂ emissions in G20 countries: Application of adaptive neuro-fuzzy inference system. *Energies* **2018**, *11*, 2771. [[CrossRef](#)]
70. Zhang, Y.; Li, X.; Zhang, Y. A novel integrated optimization model for carbon emission prediction: A case study on the group of 20. *J. Environ. Manag.* **2023**, *344*, 118422. [[CrossRef](#)]
71. Calbick, K.S.; Gunton, T. Differences among OECD countries' GHG emissions: Causes and policy implications. *Energy Policy* **2014**, *67*, 895–902. [[CrossRef](#)]
72. Pan, X.; Uddin, M.K.; Ai, B.; Pan, X.; Saima, U. Influential factors of carbon emissions intensity in OECD countries: Evidence from symbolic regression. *J. Clean. Prod.* **2019**, *220*, 1194–1201. [[CrossRef](#)]
73. Wang, W.Z.; Liu, L.C.; Liao, H.; Wei, Y.M. Impacts of urbanization on carbon emissions: An empirical analysis from OECD countries. *Energy Policy* **2021**, *151*, 112171. [[CrossRef](#)]
74. Marotta, A.; Porras-Amores, C.; Rodríguez Sánchez, A.; Villoria Sáez, P.; Masera, G. Greenhouse Gas Emissions Forecasts in Countries of the European Union by Means of a Multifactor Algorithm. *Appl. Sci.* **2023**, *13*, 8520. [[CrossRef](#)]
75. Sasaki, N.; Myint, Y.Y.; Abe, I.; Venkatappa, M. Predicting carbon emissions, emissions reductions, and carbon removal due to deforestation and plantation forests in Southeast Asia. *J. Clean. Prod.* **2021**, *312*, 127728. [[CrossRef](#)]
76. Zhang, Y.J.; Hao, J.F.; Song, J. The CO₂ emission efficiency, reduction potential and spatial clustering in China's industry: Evidence from the regional level. *Appl. Energy* **2016**, *174*, 213–223. [[CrossRef](#)]
77. Li, W.; Yang, G.; Li, X.; Sun, T.; Wang, J. Cluster analysis of the relationship between carbon dioxide emissions and economic growth. *J. Clean. Prod.* **2019**, *225*, 459–471. [[CrossRef](#)]
78. Gallo, C.; Faccilongo, N.; La Sala, P. Clustering analysis of environmental emissions: A study on Kyoto Protocol's impact on member countries. *J. Clean. Prod.* **2018**, *172*, 3685–3703. [[CrossRef](#)]
79. Lamb, W.F.; Grubb, M.; Diluiso, F.; Minx, J.C. Countries with sustained greenhouse gas emissions reductions: An analysis of trends and progress by sector. *Clim. Policy* **2022**, *22*, 1–17. [[CrossRef](#)]

80. Anwar, A.; Younis, M.; Ullah, I. Impact of urbanization and economic growth on CO₂ emission: A case of far east Asian countries. *Int. J. Environ. Res. Public Health* **2020**, *17*, 2531. [\[CrossRef\]](#)
81. Lin, T.Y.; Chiu, Y.H.; Lin, Y.N.; Chang, T.H.; Lin, P.Y. Greenhouse gas emission indicators, energy consumption efficiency, and optimal carbon emission allowance allocation of the EU countries in 2030. *Gas Sci. Eng.* **2023**, *110*, 204902. [\[CrossRef\]](#)
82. Marjanović, V.; Milovančević, M.; Mladenović, I. Prediction of GDP growth rate based on carbon dioxide (CO₂) emissions. *J. CO₂ Util.* **2016**, *16*, 212–217. [\[CrossRef\]](#)
83. Horobet, A.; Belascu, L.; Radulescu, M.; Balsalobre-Lorente, D.; Botoroga, C.A.; Negreanu, C.C. Exploring the Nexus between Greenhouse Emissions, Environmental Degradation and Green Energy in Europe: A Critique of the Environmental Kuznets Curve. *Energies (19961073)* **2024**, *17*, 5109. [\[CrossRef\]](#)
84. Wang, Q.; Yang, T.; Li, R. Does income inequality reshape the environmental Kuznets curve (EKC) hypothesis? A nonlinear panel data analysis. *Environ. Res.* **2023**, *216*, 114575. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Zhang, F.; Deng, X.; Phillips, F.; Fang, C.; Wang, C. Impacts of industrial structure and technical progress on carbon emission intensity: Evidence from 281 cities in China. *Technol. Forecast. Soc. Chang.* **2020**, *154*, 119949. [\[CrossRef\]](#)
86. Heydari, A.; Garcia, D.A.; Keynia, F.; Bisegna, F.; De Santoli, L. Renewable energies generation and carbon dioxide emission forecasting in microgrids and national grids using GRNN-GWO methodology. *Energy Procedia* **2019**, *159*, 154–159. [\[CrossRef\]](#)
87. Yu, Y.; Du, Y. Impact of technological innovation on CO₂ emissions and emissions trend prediction on ‘New Normal’ economy in China. *Atmos. Pollut. Res.* **2019**, *10*, 152–161. [\[CrossRef\]](#)
88. Ye, L.; Du, P.; Wang, S. Industrial carbon emission forecasting considering external factors based on linear and machine learning models. *J. Clean. Prod.* **2024**, *434*, 140010. [\[CrossRef\]](#)
89. Zhao, L.T.; Miao, J.; Qu, S.; Chen, X.H. A multi-factor integrated model for carbon price forecasting: Market interaction promoting carbon emission reduction. *Sci. Total Environ.* **2021**, *796*, 149110. [\[CrossRef\]](#)
90. Ameyaw, B.; Li, Y.; Annan, A.; Agyeman, J.K. West Africa’s CO₂ emissions: Investigating the economic indicators, forecasting, and proposing pathways to reduce carbon emission levels. *Environ. Sci. Pollut. Res.* **2020**, *27*, 13276–13300. [\[CrossRef\]](#)
91. Mutascu, M. CO₂ emissions in the USA: New insights based on ANN approach. *Environ. Sci. Pollut. Res.* **2022**, *29*, 68332–68356. [\[CrossRef\]](#)
92. Lee, J.; Akashi, Y.; Takaguchi, H.; Sumiyoshi, D.; Lim, J.; Ueno, T.; Maruyama, K.; Baba, Y. Forecasting model of activities of the city-level for management of CO₂ emissions applicable to various cities. *J. Environ. Manag.* **2021**, *286*, 112210. [\[CrossRef\]](#)
93. Zhong, W.; Zhai, D.; Xu, W.; Gong, W.; Yan, C.; Zhang, Y.; Qi, L. Accurate and efficient daily carbon emission forecasting based on improved ARIMA. *Appl. Energy* **2024**, *376*, 124232. [\[CrossRef\]](#)
94. Jena, P.R.; Managi, S.; Majhi, B. Forecasting the CO₂ emissions at the global level: A multilayer artificial neural network modelling. *Energies* **2021**, *14*, 6336. [\[CrossRef\]](#)
95. Jin, Y.; Sharifi, A.; Li, Z.; Chen, S.; Zeng, S.; Zhao, S. Carbon emission prediction models: A review. *Sci. Total Environ.* **2024**, *927*, 172319. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Zhou, Z.; Qiu, C.; Zhang, Y. A comparative analysis of linear regression, neural networks and random forest regression for predicting air ozone employing soft sensor models. *Sci. Rep.* **2023**, *13*, 22420. [\[CrossRef\]](#) [\[PubMed\]](#)
97. Zhao, X.; Han, M.; Ding, L.; Calin, A.C. Forecasting carbon dioxide emissions based on a hybrid of mixed data sampling regression model and back propagation neural network in the USA. *Environ. Sci. Pollut. Res.* **2018**, *25*, 2899–2910. [\[CrossRef\]](#)
98. Wang, J.; Yang, F.; Chen, K. Regional carbon emission evolution mechanism and its prediction approach: A case study of Hebei, China. *Environ. Sci. Pollut. Res.* **2019**, *26*, 28884–28897. [\[CrossRef\]](#)
99. Ashina, S.; Fujino, J.; Masui, T.; Ehara, T.; Hibino, G. A roadmap towards a low-carbon society in Japan using backcasting methodology: Feasible pathways for achieving an 80% reduction in CO₂ emissions by 2050. *Energy Policy* **2012**, *41*, 584–598. [\[CrossRef\]](#)
100. World Bank. World Development Indicators. 2020. Available online: <https://databank.worldbank.org/source/world-development-indicators> (accessed on 15 January 2024).
101. Murtagh, F.; Legendre, P. Ward’s hierarchical agglomerative clustering method: Which algorithms implement Ward’s criterion? *J. Classif.* **2014**, *31*, 274–295. [\[CrossRef\]](#)
102. Austin, P.C.; Leckie, G. The effect of number of clusters and cluster size on statistical power and Type I error rates when testing random effects variance components in multilevel linear and logistic regression models. *J. Stat. Comput. Simul.* **2018**, *88*, 3151–3163. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.