sustainability

MDPI

Article

Data-Driven Decarbonization: Machine Learning Insights into GHG Trends and Informed Policy Actions for a Sustainable Bangladesh

Md Shafiul Alam [1,*], Mohammad Shoaib Shahriar [2], Md. Ahsanul Alam [3], Waleed M. Hamanah [4], Mohammad Ali [1], Md Shafiullah [5] and Md Alamgir Hossain [6]

[1] Department of Electrical Engineering, College of Engineering, King Faisal University, Al-Ahsa 31982, Saudi Arabia; mohd.ali@kfu.edu.sa
[2] Department of Electrical Engineering, Hafar al-Batin University, Hafr Al Batin 31991, Saudi Arabia; mshoaib@uhb.edu.sa
[3] Department of Electrical and Electronic Engineering, Green University of Bangladesh, Rupganj, Narayanganj 1461, Bangladesh; ahsanul@eee.green.edu.bd
[4] ARC for Metrology, Standards, and Testing, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran 31261, Saudi Arabia; waleed.hamanah@kfupm.edu.sa
[5] Control & Instrumentation Engineering Department, Interdisciplinary Research Center for Sustainable Energy Systems, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; shafiullah@kfupm.edu.sa
[6] School of Engineering, University of Southern Queensland, Toowoomba 4350, Australia; alamgir.hossain@unisq.edu.au
* Correspondence: shafiul@kfu.edu.sa; Tel.: +966-531325436

Abstract

This work presents optimized decision tree-based ensemble machine learning models for predicting and quantifying the effects of greenhouse gas (GHG) emissions in Bangladesh. It aims to identify policy implications in response to significant environmental changes. The models analyze the emissions of CO_2, N_2O, and CH_4 from sectors including energy, industry, agriculture, and waste. We consider many parameters, including energy consumption, population, urbanization, gross domestic products, foreign direct investment, and per capita income. The data covers the period from 1971 to 2019. The model is trained using 80% of the dataset and validated using the remaining 20%. The hyperparameters, such as the number of estimators, maximum samples, maximum depth, learning rate, and minimum samples leaf, were optimized via particle swarm optimization. The models were tested, and their forecasts were extended till 2041. An examination of feature importance has identified energy consumption as a critical factor in greenhouse gas emissions, acknowledging the positive effects of clean energy in accordance with the clean development mechanism. The results demonstrate a robust model performance, with an R^2 score of approximately 0.90 for both the training and testing datasets. The bagging decision tree model showed the lowest mean squared error of 151.3453 and the lowest mean absolute percentage error of 0.1686. The findings of this study will help decision-makers understand the complex connections between socioeconomic conditions and the elements that contribute to greenhouse gas emissions. The discoveries will enable more precise monitoring of national greenhouse gas (GHG) inventories, allowing for focused efforts to mitigate climate change in Bangladesh.

Keywords: GHG emissions; climate change; emissions reduction; artificial intelligence; optimization

check for updates

Academic Editors: Alexandra Zbuchea and Mauro Romanelli

Received: 16 September 2025
Revised: 25 October 2025
Accepted: 29 October 2025
Published: 31 October 2025

Citation: Alam, M.S.; Shahriar, M.S.; Alam, M.A.; Hamanah, W.M.; Ali, M.; Shafiullah, M.; Hossain, M.A. Data-Driven Decarbonization: Machine Learning Insights into GHG Trends and Informed Policy Actions for a Sustainable Bangladesh. *Sustainability* **2025**, *17*, 9708. https://doi.org/10.3390/su17219708

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

Greenhouse gases (GHGs) are gases that warm the atmosphere by trapping some of the sun's energy. The primary greenhouse gases are carbon dioxide and water vapor, both of which are key products of hydrocarbon combustion and are among the most frequently discussed due to their high concentrations [1–3]. Approximately 73% of greenhouse gas emissions worldwide are attributed to carbon dioxide. However, other greenhouse gases, including methane, nitrous oxide, and fluorinated gases, also make substantial contributions of 19%, 5%, and 3%, respectively [4,5]. The rise in greenhouse gases directly threatens human civilization [6]. The primary drivers of GHG increases are anthropogenic activities, including the combustion of fossil fuels, deforestation, agricultural practices, and industrial processes [7,8]. GHG emissions are influenced by several factors, including population growth, economic development, power consumption, industrialization, gross domestic product (GDP), soil erosion and disturbance, and urbanization [9,10]. Across the world's continents, greenhouse gas emissions have been increasing, and the South Asian region is no exception.

Bangladesh, despite being a low emitter globally, is highly vulnerable to the impacts of climate change and to threats to food security. The agriculture sector is a major contributor to GHG emissions, driven by factors such as agricultural land expansion, crop and livestock production, energy use, and fertilizer consumption. Article [11] presents an econometric analysis of Bangladesh's greenhouse gas emissions, examining the impact of various agricultural factors. It reveals that agricultural land expansion, crop production, livestock production, fisheries production, energy use in agriculture, and fertilizer consumption increase GHG emissions in Bangladesh, while an increase in agricultural and forest land may lead to reduced GHG emissions. Article [12] focuses on carbon dioxide (CO_2) emission and its various causes, in particular, whereas [13] presents methane (CH_4) emissions in southwest Bangladesh by a Google Earth Engine-based remote sensing approach. The impact of maize cultivation on greenhouse gas emissions is addressed in [14] for Bangladesh, with total estimated emissions ranging from 1.66 to 4.09 million tons (MT) CO_2 eq. from 2015 to 2020. The paper also discusses improving the efficiency of fertilizer and water use to minimize emissions. Besides the agricultural sector, the fisheries [15] also contribute to GHG emissions, as fossil fuel sources primarily power fishing boats. Articles [16,17] discuss Bangladesh's electricity generation and its transition towards a coal-intensive mix. The paper recommended nuclear energy, electricity import, and floating solar plants for decarbonization. An exciting finding has been presented in [18] that GHG emissions depend on the type of rice cultivation. Selecting the right rice variety and water-saving irrigation method can reduce GHG emissions in Bangladesh. A two-year field experiment in Southwest China examined the effects of nutrient management practices on vegetable productivity, quality, and environmental costs [19]. Optimized nutrient management boosted pepper productivity, reduced fertilizer inputs, and improved nutrient use efficiency. The integrated knowledge and product strategy (IKPS) increased soil carbon sequestration and reduced net GHG emissions, demonstrating its potential for sustainable vegetable production.

Predicting the GHG emission pattern and its various causes is a key factor in managing this major concern. The reference [20] explores the evolution of Integrated Assessment Models (IAMs) in the climate science-policy interface, identifying five phases and their mediating roles between science and policy. It suggests expanding IAMs' scope and engaging a wider range of stakeholders. The most recent articles in the literature on GHG emission prediction have employed various machine learning and deep learning techniques. Article [21] uses a simple linear regression model, along with five other approaches, to predict the emission pattern in agricultural fields, with or without crop cover. Support

vector machine (SVM) with three kernel functions, random forest, and cubist approach of regression is used for the comparative study of this paper. The SVR approach to regression is also employed in [20,21], where particle swarm optimization (PSO) serves as a supporting tool. Article [22] modeled the relationship between global temperature and atmospheric concentrations of carbon dioxide, methane, and nitrous oxide over a 65-year timeframe using decision tree (DT), artificial neural network (ANN), random forest (RF), and linear regression. The study concluded that the ANN outperforms the other approaches. The reference [23] utilizes a Wavelet-enhanced extreme learning machine (W-EELM) with the aid of complete orthogonal decomposition (COD) to calculate the weights of ELM output layers, thereby predicting GHG emissions in short, medium, and long ranges. The paper [24] employs a data-driven intelligent time-series prediction approach to forecast greenhouse gas emissions in India, enabling policymakers to take actions that reduce emissions and achieve the targets set in the Paris Agreement. Magazzino et al. [25] used three different ML algorithms to investigate the relationship between GDP, GHG emissions, coal production, and solar and wind power generation in China, the USA, and India. The ML simulation's findings make it abundantly clear that transitioning from fossil fuels to renewable energy sources is necessary to reduce greenhouse gas emissions effectively. Mardini et al. [26] applied an adaptive neuro-fuzzy inference method and an ANN to develop multi-stage prediction models with clustered data to estimate GHG emissions for 20 countries.

As was mentioned in the discussion above, a variety of machine learning techniques have been used in numerous research projects to produce predictions about greenhouse gas emissions. However, machine learning based on decision trees (DT) demands greater concentration [27,28]. Decision trees divide the data into branches in order to extract as much information as possible. This model works well for non-linear interactions since it branches in a number of straightforward ways. This approach has several versions, and a model may occasionally incorporate multiple trees—a technique known as boosted decision trees or decision forests. The decision forest performs well on a variety of samples, but boosted decision trees—despite being more difficult to modify—can outperform the decision forest. Due to their robustness, tree-based approaches have historically been the preferred algorithm for solving classical problems. A single tree is trimmed to reduce overfitting issues, which are frequently caused by branches that prevent memory retention or memory size limitations [29].

Forecasting of Bangladesh's greenhouse gas (GHG) emissions has been conducted using various models and techniques. The Dynamic Ordinary Least Squares (DOLS) method was employed in one study [27] to investigate the impact of multiple factors on greenhouse gas emissions in the agricultural industry. Another study [30] examined the factors influencing greenhouse gas emissions, including land use, energy consumption, and population changes, using the Stochastic Impacts by Regression on Population, Affluence, and Technology (STIRPAT) model and ridge regression. Additionally, machine learning algorithms, such as linear regression and Multi-Layer Perceptron (MLP), were applied [31] to build prediction models for CO_2 emissions in Bangladesh. These models consider several factors, including energy production, the use of fossil fuels, and environmental conditions. According to the projections, several variables, including population growth, energy intensity, and agricultural practices, have an impact on CO_2 emissions. The paper [32] provides a prediction of greenhouse gas emissions from livestock in Bangladesh. The estimated emissions are expected to increase annually until 2050. Using several stepwise linear regression models, a predictive model for Bangladesh's carbon dioxide emissions is shown in [33]. The outcomes are contrasted using the Gaussian Process, Decision Table, and Random Forests.

The GHG prediction model of Saudi Arabia has been developed in [34,35]; however, hyperparameters are not optimized, and the reduction quantification is not realized. The GHG emission from the Chengdu metro in China is predicted using deep extreme learning machine learning [36]. Since the emissions from the Chengdu metro are insignificant compared to China's total emissions, a more detailed model is necessary to incorporate these total emissions. This study in [37] predicts global temperature and greenhouse gas emissions from climate change using a Recurrent Neural Network with Long-Short-Term Memory (LSTM) model. An ANN and gradient boosting-based approach is developed in [38] to analyze the impact of energy consumption on GHG emissions in the USA, China, and Europe. However, the future trend of emissions is not predicted, which is crucial for developing effective strategies and informed policy actions to mitigate climate change by understanding and addressing the sources and trends of these emissions. In the context of Bangladesh, where the application of machine learning techniques to predict greenhouse gas emissions is in its nascent stages, this study uniquely addresses the critical need for a comprehensive analysis of future emission trends, impact quantification, and policy development. The uniqueness of this study lies in the utilization of decision tree-based ensemble bagging and boosting techniques, providing an advanced methodology for forecasting emissions, measuring their influence, and drawing policy implications.

The key features of the paper are as follows:

- Three DT-regressor algorithms are implemented to solve prediction issues where modified particle swarm optimization (MPSO) is used to obtain the optimized set of hyperparameters.
- 49 years of country-reported data are used for Bangladesh to predict the GHG emission behavior up to the year 2041, where the training and testing have been validated through scatter plot graphs.
- Permutation and SHapley Additive exPlanations (SHAP) analysis are used to differentiate between the input features of the algorithms and rank them according to their importance.
- GHG Reduction quantification is realized, along with the predicted rising pattern.
- Informed policy actions are suggested for GHG reductions based on the findings.

The remaining article has the following structure. Materials and methods are provided in Section 2. Performance indices are documented in Section 3. Results and discussions are provided in Section 4. Finally, Section 5 summarizes this work and provides direction for future research.

2. Materials and Methods

2.1. Data Descriptions

In order to generate a comprehensive set of greenhouse gas emission trajectories for every nation and Kyoto gas covering the years 1750 to 2019, the UNFCCC member states, as well as the majority of non-UNFCCC territories, combined multiple published datasets. The main Intergovernmental Panel on Climate Change (IPCC) 2006 categories are resolved by the data. For CO_2, CH_4, and N_2O, there are subsector statistics for energy, industrial processes, and agriculture. PRIMAP-hist dataset compiles both country-reported and third-party data [39]. However, in this research, only country-reported data are used. The World Bank repository contains various sources of data, including the United Nations Population Division's World Population Prospects: 2019 Revision, census reports and other statistical publications from national statistical offices, demographic statistics from Eurostat, population and vital statistics reports from the United Nations Statistical Division (in various years), the U.S. Census Bureau's International Database, and the Secretariat of the Pacific Community's Statistics and Demography Programme. Each

country's population statistics are available from 1960 to 2021 [40]. The socioeconomic and energy-related indicators like population, GDP, per capita income, energy use, and urban development were collected from official country-reported and institutional datasets from different sources like UN, World Bank, and IEA [41,42]. However, as the GHG data is only accessible until the year 2019, the population data from the same year is utilized for the country in the building of the model. In summary, it is acknowledged that such sources may contain inherent reporting inconsistencies, time lags, or incomplete sectoral coverage. Previous studies [34] have also relied on similar national datasets, emphasizing their significance for maintaining coherence with governmental climate policy targets and emission accounting frameworks. Therefore, this research follows that established approach to maintain comparability and contextual accuracy in evaluating the country's emission trajectories and model predictions.

Table 1 presents a thorough summary of important economic and demographic information for Bangladesh. The population varies from 68.38 million to 165.52 million, with an average of 117.25 million. The GDP exhibits substantial variation, ranging from a minimum of USD 6.29 billion to a maximum of USD 351.00 billion. The countries also display significant disparities in net national income and urbanization rates. Foreign Direct Investment (FDI) fluctuates between negative USD 0.01 billion to positive USD 2.83 billion, with an average of USD 0.54 billion. The Gross National Income (GNI) per capita range varies from USD 81.14 to USD 2197.90, indicating substantial economic variation. The per capita energy consumption and greenhouse gas emissions are quantified in kilograms. The average energy consumption per capita is 145.75 kg, while the average greenhouse gas emissions are 78.86 Mt CO_2 eq. The data emphasizes the significant differences in economic and environmental aspects, as seen by the kurtosis and skewness values, which describe the distribution features of each variable. The scatter plot of the whole dataset is provided in Figure 1.

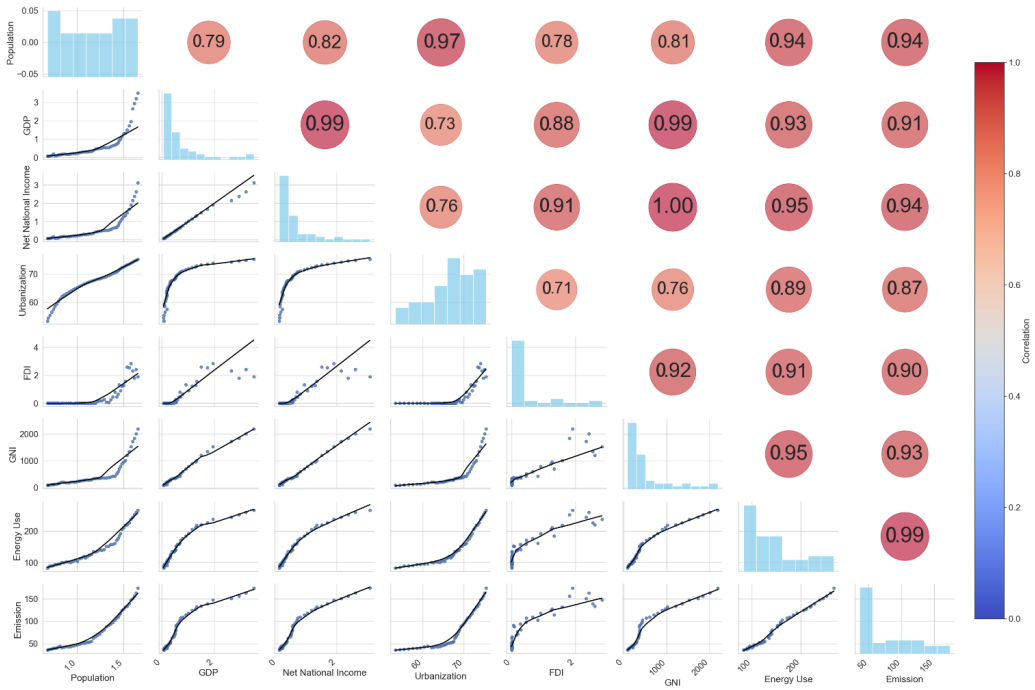


Figure 1. Scatter plot of Bangladesh dataset on GHG emissions.

Table 1. Statistical summary of economic and GHG dataset.

	Population (Million)	GDP (Billion USD)	Net National Income (Billion USD)	Urbanization Rate	FDI (Current Billion USD)	GNI (per Capita Current USD)	Energy Use (kg per Capita)	GHG (Mt CO_2 eq.)
count	49	49	49	49	49	49	49	49
mean	117.25	72.32	66.65	66.89	0.54	531.70	145.75	78.86
std	30.11	84.01	72.85	5.75	0.86	541.99	53.33	41.07
min	68.38	6.29	7.28	53.36	−0.01	81.14	83.16	35.88
25%	91.05	19.45	17.91	63.60	0.00	200.30	101.87	43.69
50%	117.79	37.94	36.01	67.78	0.01	331.26	133.42	62.04
75%	144.14	79.61	77.48	71.28	0.81	599.48	177.02	109.18
max	165.52	351.00	312.00	75.35	2.83	2197.90	266.88	173.82
kurtosis	−1.30	3.56	2.70	−0.30	0.99	2.20	−0.46	−0.72
skewness	−0.06	2.02	1.79	−0.63	1.52	1.76	0.84	0.77
range	97.14	344.71	304.72	21.99	2.84	2116.76	183.72	137.94

2.2. Model Descriptions

2.2.1. Bagging Regressor

Bagging, also known as Bootstrap Aggregating, is a powerful ensemble learning method intended to improve the reliability and efficiency of prediction models. In regression analysis, Bagging Regressor emerges as a prominent methodology, harnessing the collective intelligence of multiple weak learners to create a more accurate and stable predictive model.

The fundamental concept behind the Bagging Regressor lies in the generation of diverse and independently trained sub-models through the application of bootstrapping [43]. Bootstrapping encompasses drawing random samples with replacements from the original dataset, thereby creating multiple subsets for training individual models. Each subset is utilized to train a separate regressor, and their predictions are aggregated to form the final output. Figure 2 shows the architecture of the Bagging Regressor.

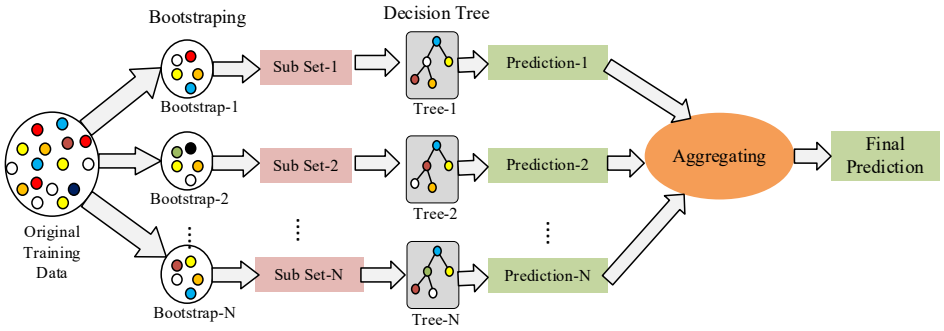


Figure 2. The architecture of Bagging Regressor.

One of the key strengths of Bagging Regressor is its ability to mitigate overfitting by introducing variability in the training process. By training on different subsets of data, the individual models capture unique patterns and nuances present in the dataset. The ensemble effect is achieved through averaging or voting, where the diverse predictions of individual models contribute to a more robust and generalized prediction [44].

2.2.2. Adaboost Regressor

Adaboost, short for Adaptive Boosting, is a popular boosting algorithm that employs a distinctive strategy to combine the strength of weak learners in a sequential and adaptive fashion. The fundamental idea behind boosting is to sequentially train weak learners, with each subsequent model focusing on the errors made by its predecessors. Adaboost is a

widely used boosting algorithm for both classification and regression tasks [45]. It works by giving more weight to misclassified samples, forcing the subsequent weak learners to focus on the previously misclassified instances.

The schematic diagram of Adaboost regression is shown in Figure 3 [46]. The algorithm initiates with a basic model and iteratively introduces subsequent weak models, each concentrating on rectifying the errors of its forerunner. This sequential refinement process sets AdaBoost apart, allowing it to adapt and evolve its predictive ability with each iteration.

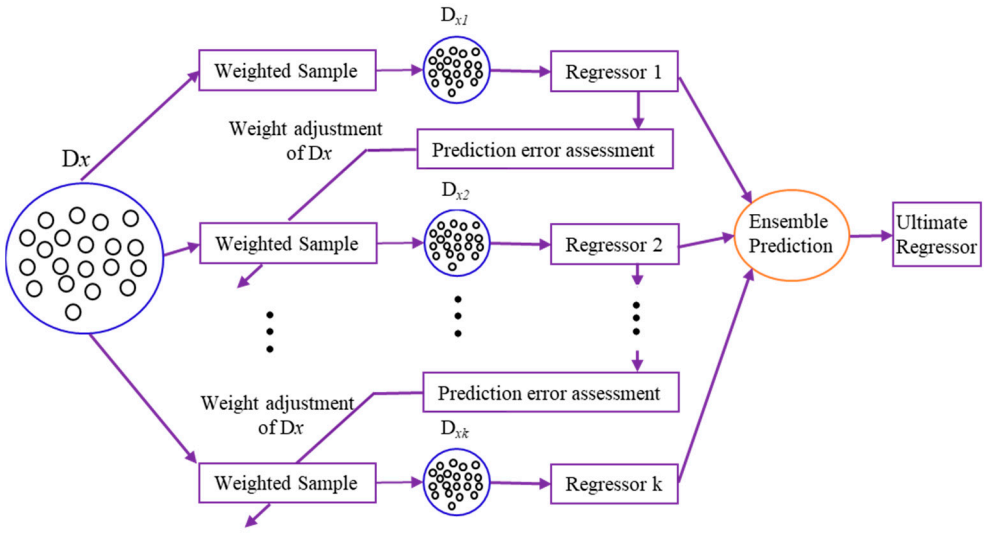


Figure 3. Schematic diagram of Adaboost regression.

A hallmark feature of AdaBoost is its emphasis on the weighted contribution of each weak learner. Instances that are inadequately predicted are assigned higher weights, directing the algorithm's attention towards more accurate predictions.

2.2.3. Gradient Boosting Regressor

Gradient Boosting Regressor (GBR) is a powerful machine learning algorithm that has gained popularity for its ability to produce highly accurate predictions in regression problems. It belongs to the family of ensemble methods, which combine the predictions of multiple weak learners to create a strong learner. Developed as an extension of the Gradient Boosting Machine (GBM) algorithm [47], GBR sequentially trains weak learners, usually decision trees, and correct the errors made by their predecessors. The term "gradient boosting" refers to the technique of minimizing the residuals or errors of the previous models in the ensemble [48]. As shown in Figure 4, the core components of a Gradient Boosting Regressor are base estimators, loss function, gradient descent, learning rate, and tree pruning.

Gradient Boosting Regression (GBR) employs weak learners, usually decision trees with limited depths, as its fundamental building blocks, leveraging their combination in an ensemble to enhance predictive accuracy. During training, the method minimizes a specified loss function, such as Mean Squared Error or Huber loss, to align predicted values with actual outcomes. Gradient descent optimization is employed iteratively, where each step involves calculating gradients and training weak learners on residuals to reduce prediction errors progressively. A crucial hyperparameter, the learning rate, controls the optimization step size, acting as a regularization term to balance model complexity and stability. Additionally, tree pruning is implemented to prevent overfitting, limiting the depth of decision trees and prioritizing the capture of relevant data patterns while avoiding

noise memorization. The provided flow diagram in Figure 4 illustrates the sequential process of the gradient boosting machine method.

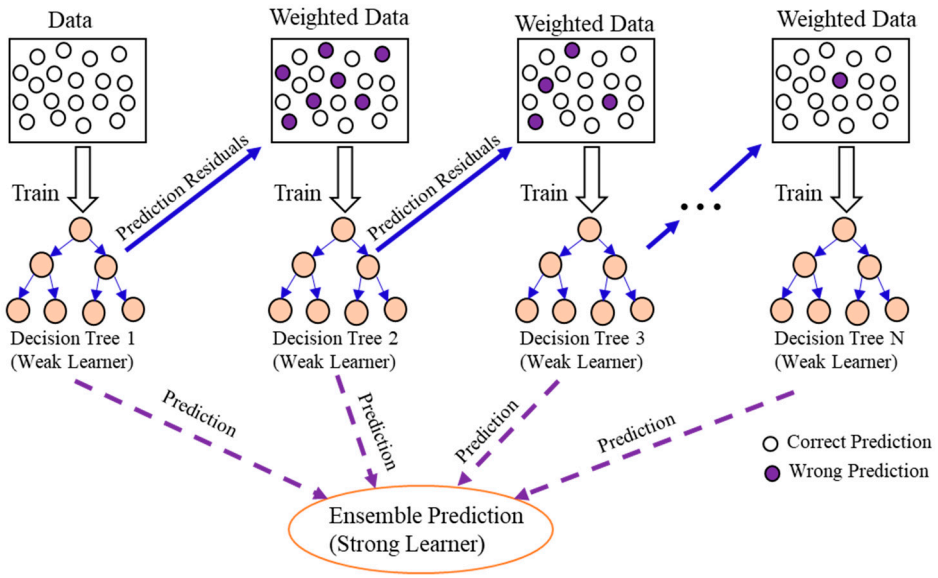


Figure 4. Flow diagram of the gradient boosting method.

2.2.4. Modified Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a popular metaheuristic optimization technique inspired by the social behavior of bird flocks. It has been extensively used in various optimization problems due to its simplicity and effectiveness. However, the basic PSO algorithm, introduced by Kennedy and Eberhart in 1995, has some limitations and drawbacks in handling complex optimization tasks [49]. Several modifications and enhancements have been proposed to address these limitations.

In PSO, a population of candidate solutions, known as particles, moves through the search space to find the optimal solution. Each particle adjusts its position based on its own experience (local best) and the experiences of its neighbors (global best). The movement of particles in PSO is governed by velocity and position updates according to the following equations:

$$v_{ij}(t+1) = wv_{ij}(t) + c_1 r_1 \left(p_{ij} - x_{ij}(t)\right) + c_2 r_2 \left(p_{gj} - x_{ij}(t)\right) \quad (1)$$

$$x_{ij}(t+1) = x_{ij}(t) + v_{ij}(t+1) \quad (2)$$

where

$v_{ij}(t)$ is the velocity of particle i in dimension j at time t.
$x_{ij}(t)$ is the position of particle i in dimension j at time t .
p_{ij} is the personal best position of particle i in dimension j .
p_{gj} is the globa best position of particle i in dimension j .
c_1 and c_2 are acceleration coefficients.
r_1 and r_2 are random numbers in the range [0, 1].
w is the inertia weight.

While traditional PSO has shown promising results in many applications, it suffers from issues like premature convergence, poor exploration-exploitation in the search space, sensitivity to parameter settings, and inability to handle many constraints. These limitations have led to the development of modified versions of PSO (MPSO) aiming to enhance its

convergence speed, exploration-exploitation balance, and performance on different types of optimization problems. Among some of the key modifications in MPSO are [50]:

(i) Constriction Factor: Clerc and Kennedy [51] introduced the constriction factor K to adjust the velocity of particles, as shown in Equation (3). This modification aims to improve the convergence properties of PSO.

$$v_{ij}(t+1) = K\left[v_{ij}(t)\; c_1 r_1 \left(p_{ij} - x_{ij}(t)\right)\; c_2 r_2 \left(p_{gj} - x_{ij}(t)\right)\right] \tag{3}$$

where the constriction factor K is given by Equation (4).

$$K = \frac{k}{\left|2 - \varphi - \sqrt{\varphi^2 - 4\varphi}\right|} \tag{4}$$

where $k = 2,$ and $\varphi = c_1 + c_2,$ and $\varphi > 4$.

A time-dependent linearly decreasing K performs better than the fixed one [52].

(ii) Control Parameter Adjustment: In the context of PSO algorithm, control parameters refer to inertia weight (w) and acceleration factors (c_1 and c_2). One of the critical modifications in MPSO involves the adjustment of these parameters. In standard PSO, these parameters control the balance between exploration and exploitation. Over the last few decades, many strategies have been proposed for the adjustment of these parameters [49,53]. Some of these are listed in Figure 5. In adaptive strategies, these parameters are adjusted dynamically during the optimization process. This adaptation helps in achieving a better balance between global and local search, leading to improved convergence and exploration capabilities.

(iii) Constraint Handling: Standard PSO struggles with optimization problems involving constraints. MPSO incorporates mechanisms to handle constraints more effectively. Constraint-handling techniques, such as penalty functions or repair operators, are integrated into MPSO to ensure that solutions generated during the optimization process adhere to the problem constraints. This modification broadens the applicability of MPSO to a wider range of real-world optimization problems.

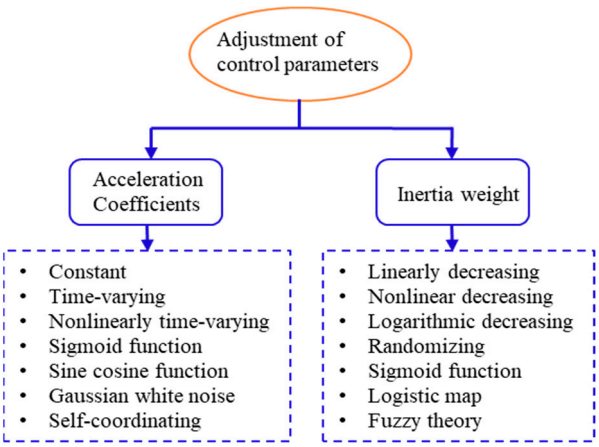


Figure 5. PSO variants with modified control parameters.

MPSO is specifically chosen over simpler optimization techniques (such as Grid Search or Random Search) because of its metaheuristic nature and adaptive parameters (w, c_1, c_2), which offer a superior balance between exploration and exploitation in complex, high-dimensional search spaces, a characteristic of ensemble machine learning hyperparameter tuning.

2.2.5. Methodology Flowchart

Our approach incorporates bagging and boosting ensemble models into the framework, considering MPSO into the hyperparameters tuning loop as shown in Figure 6. The framework starts economic data processing, including population, GDP, net national income, urbanization, FDI, GNI, and energy use. The country reported GHG data from several sectors, including industrial process and product use, energy, waste, and agriculture sectors, which are processed to obtain equivalent CO_2 emissions. The data, starting from the year 1971 up to the year 2019, is split into training and test sets. This study utilizes an 80/20 data split for model validation; however, advanced methodologies such as k-fold cross-validation, bootstrapping, and uncertainty analysis are acknowledged in the literature as critical strategies for augmenting model robustness. The reference [54] examines the impact of greenhouse gas concentrations on trapped solar radiation and the subsequent effects on temperature trends. It uses several machine learning models and k-fold cross-validation to make predictions more reliable and models stronger. The reference [55] discusses how phase change materials can be used in buildings and how machine learning can improve system design, operational control, and performance prediction. It also gives an overview of how to do sensitivity and uncertainty analysis in climate prediction to help make energy solutions that are more reliable and adaptable. Other advanced model validation techniques are reported in [56,57] for climate predictions. In the next step of this study, bagging and boosting models are trained, and the important hyperparameters are well-tuned, applying MPSO for better prediction. Optimized hyperparameters include the number of estimators, maximum samples, maximum depth, learning rate, and minimum samples leaf. After the models have been trained with the optimized parameters, they are saved using a dump. In Python 3.12, to serialize and store data for later use pickle module is utilized, as demonstrated by dumping a dictionary into a file. Afterward, the trained models are used to analyze the importance of the features. The feature importance analysis conducted on machine learning models for GHG emission prediction in Bangladesh plays a pivotal role in enhancing our understanding of the factors influencing emissions. By scrutinizing variables such as energy consumption, population, urbanization, gross domestic product, foreign direct investment, and per capita income, this analysis not only provides valuable insights into the intricate relationships within the dataset but also identifies the most significant contributors to greenhouse gas emissions. The performance of models is tested using several statistical measurements, including R^2 score, mean squared error, mean absolute error, and mean absolute percentage error. Graphical performance analyses, such as radar plot, scatter plot, bar plot, time-series plot, are also conducted. The reduction is quantified and plotted with future predictions to understand the GHG emission trends. It is worth mentioning that the analysis employs a linear projection methodology owing to data and scope constraints, thereby limiting the incorporation of multi-scenario modeling or Integrated Assessment Models (IAMs). This method makes future paths easier to understand, but it does not take into account uncertainties in policy, technology, or the economy. Similarly, due to computational constraints, a systematic performance comparison of MPSO against other established hyperparameter tuning methods, such as Random Search or Bayesian Optimization [58], was not conducted in this study. This methodological constraint necessitates further investigation in future research to validate the optimization gains afforded by MPSO comprehensively. The limitation underscores the necessity of integrating scenario-based projections and uncertainty simulations in forthcoming research to augment forecasting robustness.

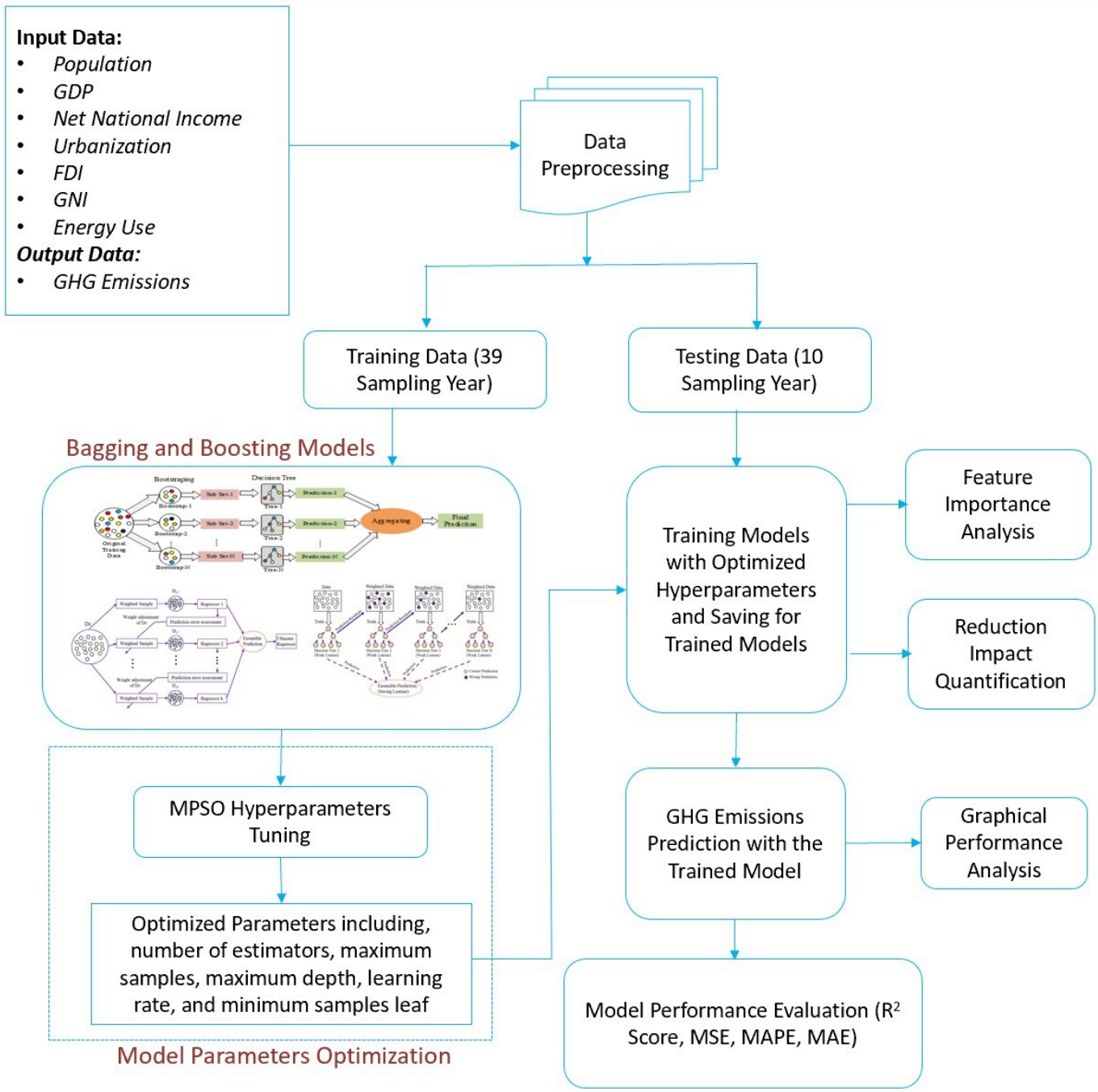


Figure 6. Flowchart of the proposed approach.

An extensive analysis of vehicle emissions and their impact on policy analysis in South Korea has been presented in [59]. The reference [60] presents a sustainable policy analysis of GHG emissions for the UK, whereas [61] presents a similar comparative analysis across multiple countries. Several works are also conducted in Bangladesh, including [62], which discusses a sustainable waste management policy to reduce greenhouse gas emissions. This study offers data-driven insights that can facilitate informed policy decisions in Bangladesh, as implicitly mentioned in Figure 6. By figuring out what causes greenhouse gas emissions with feature importance analysis, especially how much energy is used, policymakers can see that they should focus on deploying renewable energy, improving energy efficiency, and developing specific plans to reduce carbon emissions that align with international climate goals [63]. The results help link socioeconomic trends to emission

patterns, thereby making national climate policies more effective and focused. However, the analysis relies on historical data up to 2019 and excludes scenario-based modeling, policy shifts, or technological advancements, potentially undermining the durability of long-term policy recommendations. Furthermore, this study's methodological limitation is its exclusive reliance on three ensemble models, such as Bagging, AdaBoost, and Gradient Boosting, neglecting other advanced methodologies such as deep neural networks, hybrid models, or complex time series forecasting techniques (e.g., LSTM, Prophet). This limits the investigation of potentially more potent and resilient predictive methodologies recorded in the literature.

3. Performance Matrices

This study makes use of four different statistical metrics in order to evaluate the performance of the developed models. Some examples of these include the mean absolute error (MAE), the mean squared error (MSE), the mean absolute percentage error (MAPE), and the R^2 score. By dividing the MAE of the actual data by the MAE of the forecast errors, the MAPE can be calculated. There is a correlation between lowering the values of MAPE, MAE, and MSE and improved model performance. At the other end of the spectrum, a larger R^2 score indicates superior performance. If a regression model's independent variables are able to explain the variance in the dependent variable that is being evaluated, then the R^2 score reflects the extent to which this is possible.

$$R^2 = 1 - \frac{SSR}{SST} \tag{5}$$

where *SST* represents total sum of squares and *SSR* represents sum of squared regression as captured in Equation (5).

The MAPE is defined by the following equation.

$$MAPE = \frac{\sum_{i=1}^{N} |dO_i - dF_i|}{dO_i} \times 100\% \tag{6}$$

In Equation (6), dF_i is the prediction data, dO_i is the observed data, and N is the number of samples.

If the predicted and observed values are identical, the MSE equals 0, indicating no model error. MSE is often calculated using the formula below, and its value increases as the model error does.

$$MSE = \frac{\sum_{i=1}^{N} (dO_i - dF_i)^2}{N} \tag{7}$$

where dO_i is the ith observed value and dF_i is the corresponding predicted value as shown in Equation (7).

The MAE measures the average forecasting error size without considering its direction. It assesses precision for variables that are continuous and is given by the equation below.

$$MAE = \frac{\sum_{i=1}^{N} |dA_i - dF_i|}{N} \tag{8}$$

In Equation (8), *dAi* is the corresponding actual value while others are as defined above.

4. Results and Discussion

This section presents the results of using several decision tree machine learning algorithms, including the bagged decision tree, boosted decision tree, and gradient boosted decision tree, to forecast Bangladesh's greenhouse gas emissions. The efficacy of the algorithms will also be checked and verified through standard tests. In addition, the features used in machine learning models will be categorized based on their importance through the Shapley explanation model. The results are presented below under various subsections.

4.1. Hyperparameters for Models

Three different machine learning algorithms are used in this research to predict greenhouse gas emissions. The maximum and minimum range of different hyperparameters are provided in Table 2.

Table 2. The Range of hyperparameters for several models.

Algorithms	Hyperparameters Name	Minimum	Maximum
Bagging	n_estimators	50	250
	max_samples	0.1	1.0
	max_features	0.1	1.0
Gradient Boost	learning_rate	0.01	0.2
	n_estimators	50	250
	max_depth	2	20
	min_samples_split	2	20
	min_samples_leaf	1	10
Adaboost	learning_rate	0.01	0.2
	n_estimators	50	250

The population size of 100, acceleration coefficients c_1 and c_2 of 0.2, and random numbers r_1 and r_2 uniformly distributed over [0, 1] are used in this study for the MPSO algorithm.

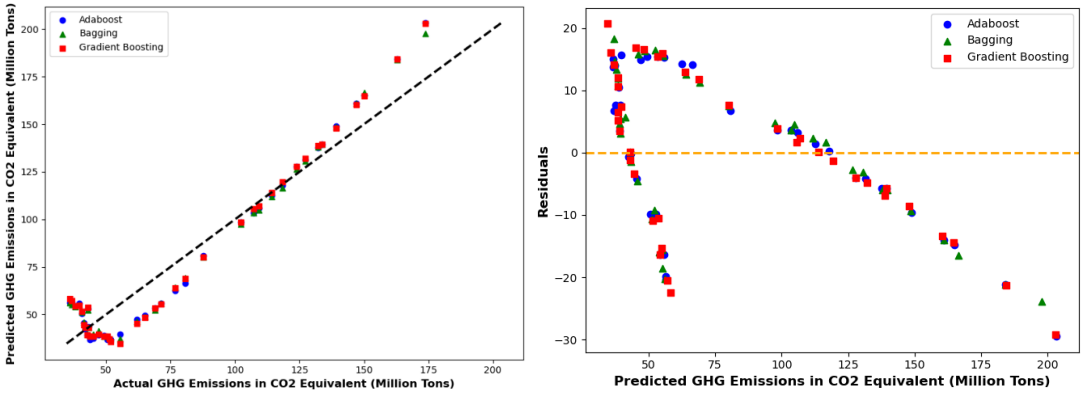
It is very crucial to select the hyperparameters of these algorithms correctly, as the performance of the methods depends on their correct selection. Therefore, the well-known metaheuristic optimization tool, modified particle swarm optimization (MPSO), is employed to obtain the optimal combination of parameters. The optimized parameters for different algorithms to conduct prediction and analysis tasks are presented in Table 3. It has been observed that the R^2 score reached close to unity for all the regression models, which clearly reflects the efficacy of the algorithms. The bagging regressor requires approximately 40 iterations to provide the optimal set of solutions. In contrast, the gradient boosting and AdaBoost regression algorithms require around 25 and 55 iterations, respectively, to achieve the maximum R^2 score.

Table 3. Optimized hyperparameters for several models.

Algorithms	Hyperparameters
Bagging	n_estimators = 61, max_samples = 0.877, max_features = 0.566
Gradient Boost	max_depth = 10, min_samples_split = 11, min_samples_leaf = 4, n_estimators = 161, learning_rate = 0.179
Adaboost	n_estimators = 62, learning_rate = 0.171

4.2. Model Fitting Results

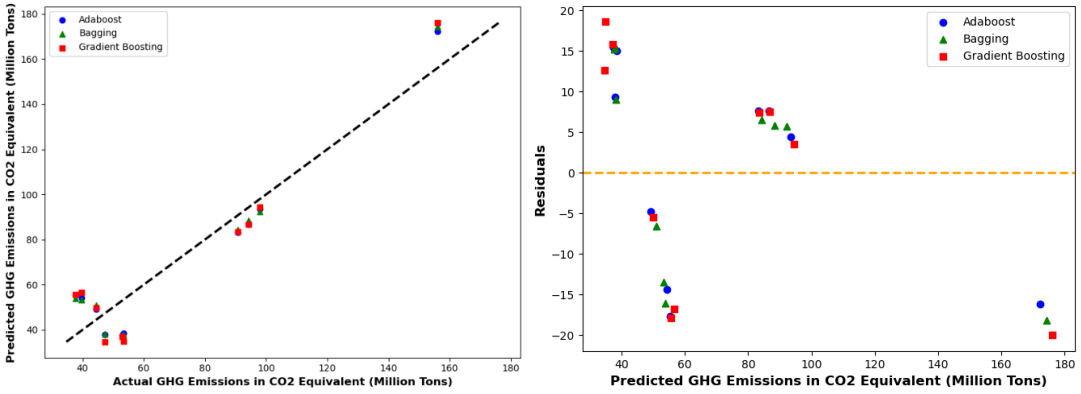
Figure 7a,b depict the training efficacy of three ensemble models, such as Adaboost, Bagging, and Gradient Boosting, in forecasting GHG emissions in Bangladesh. In the scatter plot (Figure 7a), all models exhibit robust concordance with the actual emissions, particularly between 75 and 150 million tons, where data points are positioned near the ideal line. Bagging and Gradient Boosting predictions closely align with the actual trend, however Adaboost tends to underpredict at elevated emission levels (exceeding 175 million tons). In the residual plot (Figure 7b), the residuals are predominantly clustered around zero for most predictions, indicating negligible errors. Adaboost and Bagging exhibit significant positive residuals at lower emission levels (below 75 million tons), indicating underprediction, but negative residuals emerge at higher values, implying overprediction.



(**a**) Scatter plot for training data (**b**) Scatter plot for training data

Figure 7. Performance evaluation of the bagging and boosting algorithms on training dataset.

Figure 8a,b present identical assessments for the testing dataset. The scatter plot (Figure 8a) reaffirms the predictive efficacy of the models, as the majority of forecasts correspond closely with actual values. Gradient Boosting exhibits marginally superior performance, particularly within the range of 60 to 140 million tons. The residual plot (Figure 8b) demonstrates that all models have low residuals, signifying effective generalization.



(**a**) Scatter plot for testing data (**b**) Scatter plot for testing data

Figure 8. Performance evaluation of the bagging and boosting algorithms on testing dataset.

Figure 9 shows the performance of the ML models on the training dataset in predicting yearly GHG emissions. The observed data indicates that, for the majority of data points, the predicted and actual data are highly similar, ensuring reliable predictions from the proposed algorithm. The predicted and actual emission graphs follow a similar trend in data points 8–10. Similarly, the prediction error is very low around data point 30 where the

estimated curve coincides with the actual curve. The data points with the largest prediction errors are 1, 17, and 39, where the prediction graph is not fully matched with the actual emission plot.

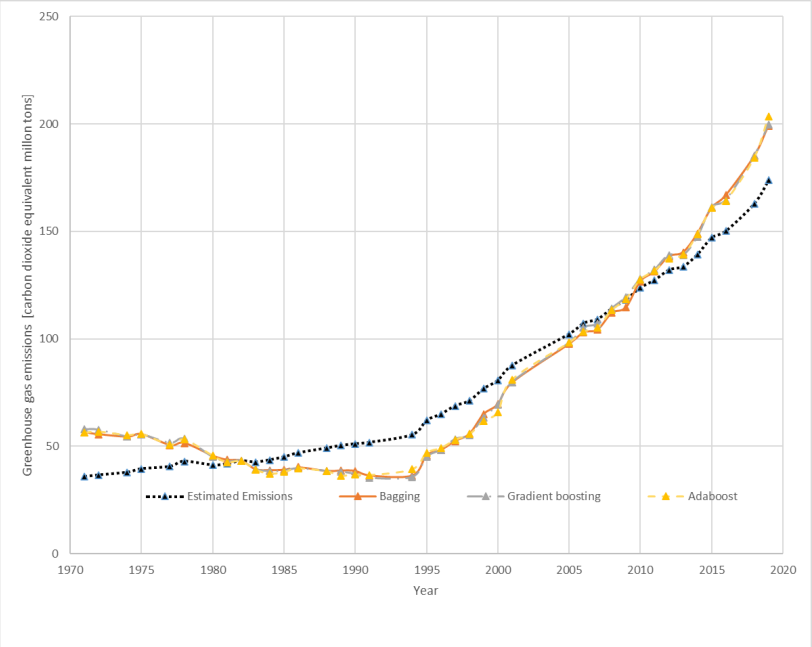


Figure 9. Performance of the models on the training dataset.

In the case of testing dataset performance curves, presented in Figure 10, data 1, 5 and 10 show the maximum error, whereas the actual and estimated curves overlap each other between data points 3 and 4. A similar phenomenon is also observed between points 9 and 10. However, the estimated curves for all three ML algorithms are seen very close to the actual curve, signifying the efficacy of the models on the training and testing datasets.

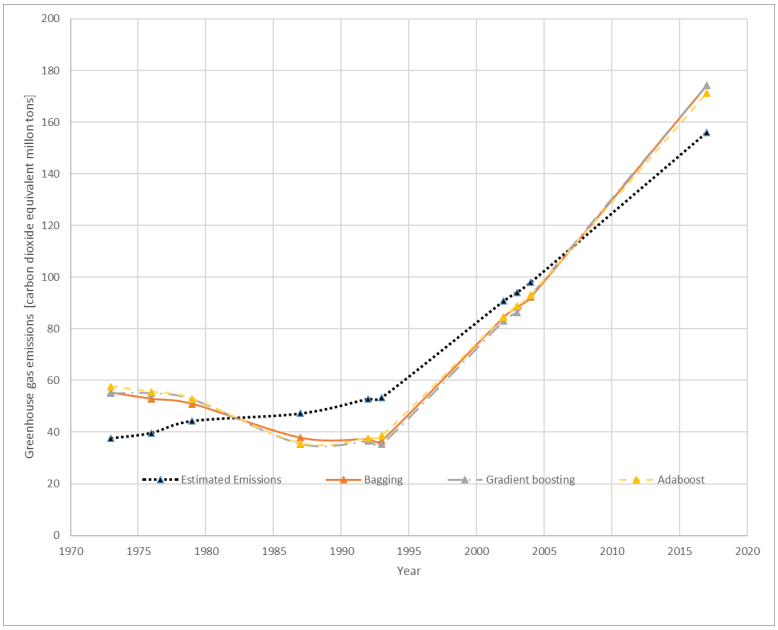


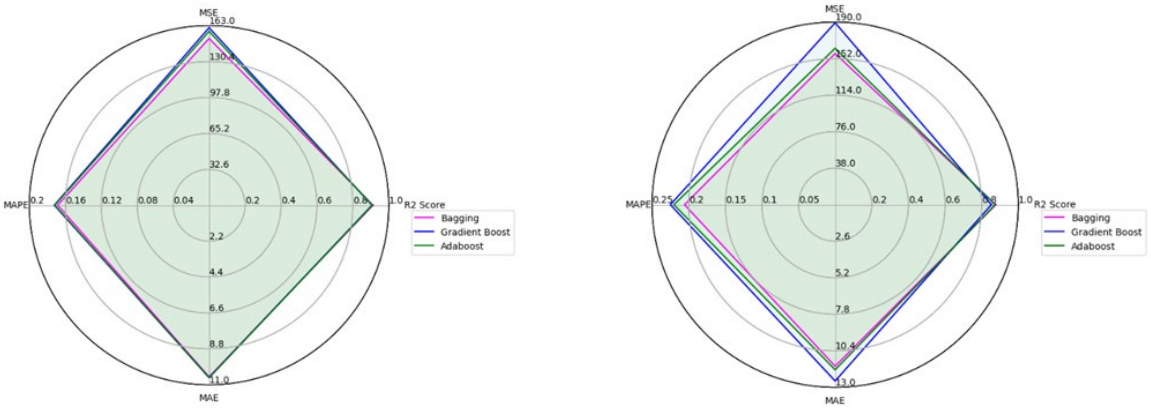
Figure 10. Performance of the models on the testing dataset.

4.3. Performance Measurement Through Statistical Measures

The machine learning algorithms' prediction results are tested and verified through various indices, including R^2 score, MSE, MAPE, MAE, RMSE, Nash–Sutcliffe efficiency coefficient (NSE), and so on. A significant limitation of this study is that model validation relied on a singular 80/20 data split instead of more sophisticated methodologies like k-fold cross-validation or uncertainty analysis, potentially constraining the generalizability of the findings. Three ensemble model's statistical performance metrics for training and testing datasets are displayed in Table 4. For both the training and testing dataset, the bagged decision tree exhibits the best R^2 performance, with the values of 0.9124 and 0.8780 for training and testing datasets, respectively. The mean squared error for the bagging algorithm is found to be 151.34 and 157.27 for training and testing, respectively. This indicates the optimal performance by the mentioned algorithm, which also displays the values of other error parameters: MAPE and MAE. For both the training and testing data, the error indicator values for bagging decision tree estimator outperforms the other two approaches: gradient boost and adaboost. However, the three error indicator values for the other two algorithms are very close and slightly higher than the bagging estimator values. The radar plots of the performance indices for training and testing data are visualized in Figures 11a and 11b, respectively.

Table 4. Statistical performance of bagging and boosting models.

Model	Training/Testing	R^2 Score	MSE	MAPE	MAE	RMSE	NSE
Bagging	Training	0.9124	151.3453	0.1686	10.5129	12.0358	0.9162
	Testing	0.8780	157.2719	0.2058	11.4975	12.1430	0.8856
Gradient Boost	Training	0.9067	161.1563	0.1724	10.5611	12.6923	0.9068
	Testing	0.8532	189.2605	0.2251	12.5496	13.7222	0.8539
Adaboost	Training	0.9086	157.8853	0.1718	10.5689	12.5190	0.9093
	Testing	0.8737	162.8069	0.2203	11.7517	12.1187	0.8861



(**a**) Training dataset (**b**) Testing dataset

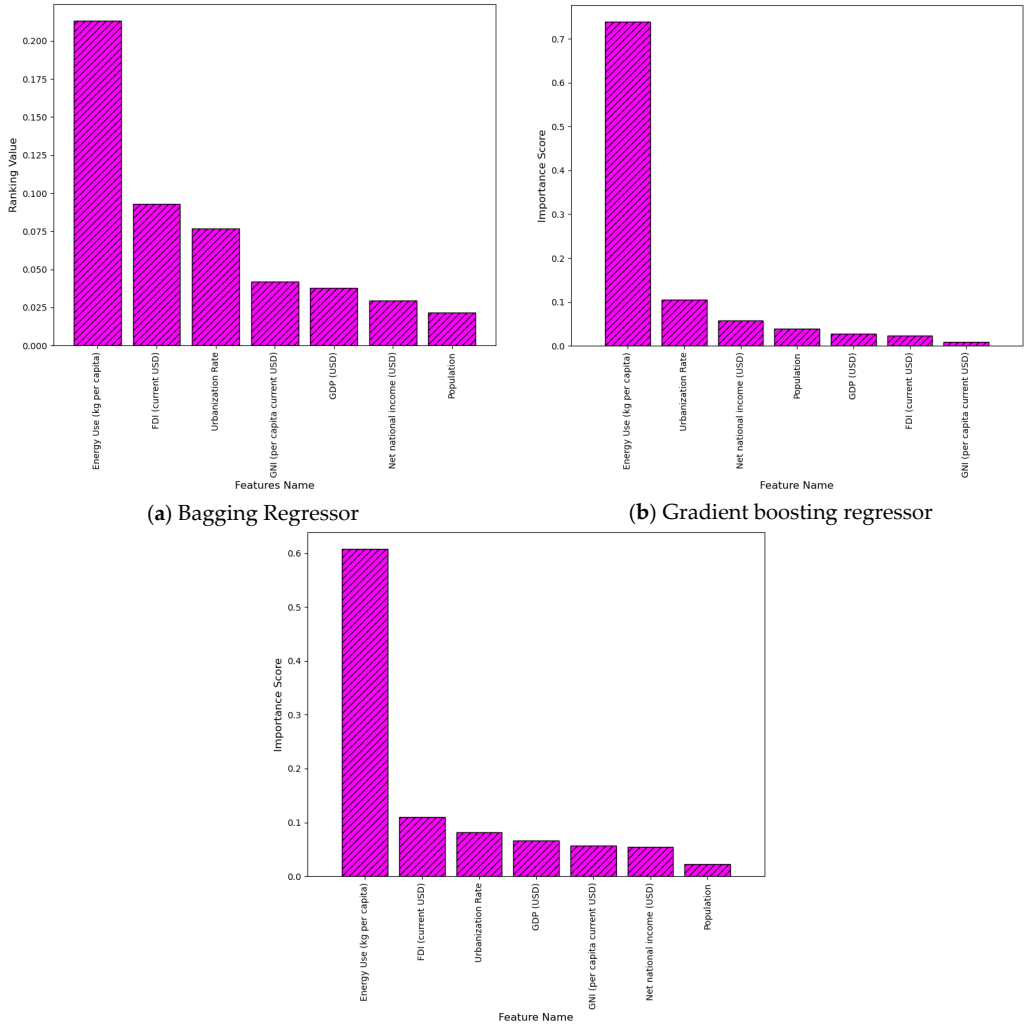
Figure 11. Radar plot of statistical performance.

4.4. Categorization of the Features by Shapely Explanation

As discussed above, various features were used in this research to create the dataset for training and testing. However, all of them are not of the same importance. Some of the features might have a higher impact on the regression model over others. This can be found from the feature importance analyses including SHAP and permutation, which are used to explain the output of machine learning models by assigning credit for a model's prediction

to each feature or feature value [64]. These are influential instruments in comprehending how machine learning model's function. They serve to clarify the predictions generated by diverse types of models, such as linear regression, generalized additive regression, non-linear boosted tree models, linear logistic regression and non-linear boosted tree logistic regression.

Figure 12 presents the result of the permutation test for the machine learning models used in this research. Among the features used, the impact of 'energy use (kg per capita)' is seen as the most significant among all, with the highest value. This phenomenon has been observed for all three developed regressors. The remaining features can be arranged as per the ascending order of the permutation test score of the bagging regressor (Figure 12a) as follows: FDI (current USD), urbanization rate, net national income (USD), GDP (USD), population and GNI (per capita current USD). For the gradient boosting regressor, the order is different, and the urbanization rate becomes the second most impactful feature, as shown in Figure 12b. Meanwhile, the GNI (per capita current USD) resulted in the second-highest value for the Adaboost estimator, as reflected in Figure 12c.



(**a**) Bagging Regressor (**b**) Gradient boosting regressor

(**c**) Adaboost regressor

Figure 12. Importance of several features in predicting GHG emissions.

Figure 13 displays a Taylor diagram assessing the efficacy of the Bagging Regression model in forecasting greenhouse gas (GHG) emissions in Bangladesh. The observation point, depicted in red, closely matches the predicted model arc, demonstrating a strong

correlation, a low root mean square deviation (RMSD), and a consistent standard deviation, thereby affirming the model's resilience.

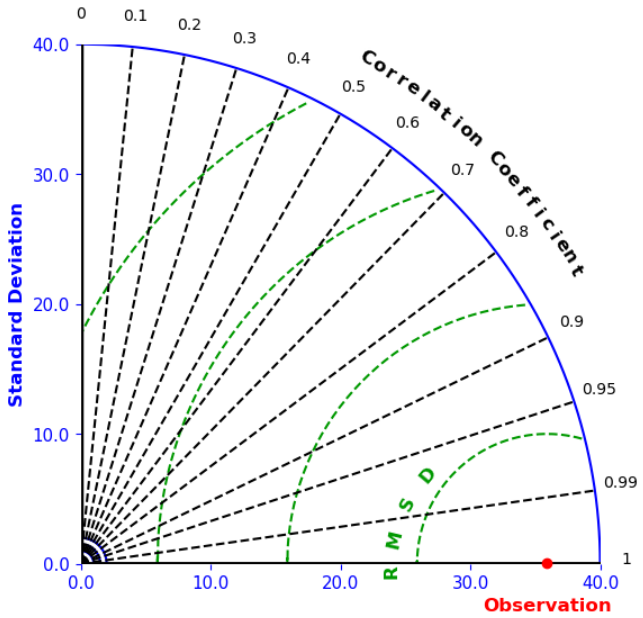


Figure 13. Taylor diagram showing high agreement between predicted and observed GHG emissions using Bagging Regression.

Figure 14 illustrates the SHAP summary plot for the same model, highlighting the contribution of individual features to the model's predictions. Notably, GDP appears as the most influential feature in SHAP analysis, indicating its consistent and strong contribution across instances. However, this differs from the permutation importance results, where energy use per capita was ranked highest. This mismatch occurs because SHAP values indicate local interpretability and interaction effects on individual predictions, whereas permutation significance assesses the global influence on model accuracy when each feature is randomized. Collectively, these metrics provide additional insights into the model's reliability and interpretability. Previous research has shown that distinct yet complementary insights could be obtained from different machine learning approaches [65]. The studies presented in [66] showing several ensemble models with different features have the highest impact on predicting traffic crash severity.

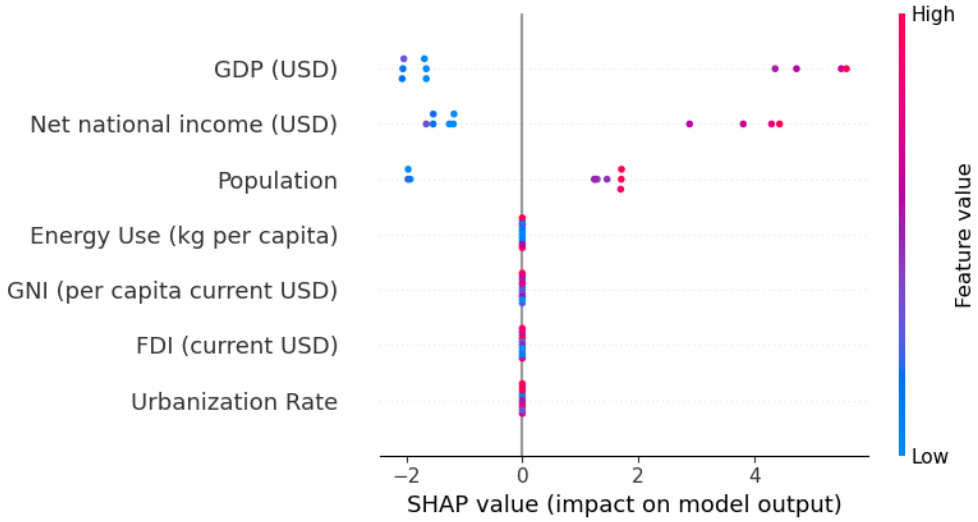


Figure 14. SHAP summary plot.

4.5. Prediction of GHG Emission in the Future

Based on the GHG emission data of 40 years, from 1971 to 2019, the emission pattern for the next 22 years has been predicted by three regression models. Using the input variables from 2020 to 2041, the ensemble model has been developed to predict greenhouse gas emissions. Figure 15 illustrates the increasing trend in greenhouse gas emissions across all models, which include the bagging model, Ada boosted model, and gradient boosted model. The largest GHG emissions, about 205 million CO_2 equivalent tons, are predicted by the Adaboost regression model to occur in 2041. The gradient-boosted model, out of all the models, projects Bangladesh to have the lowest GHG emissions in 2041, around 200 million CO_2 equivalent tons. The predicted curves are linearly rising in nature for all three algorithms, where the curve for bagging regressor falls in between the other two. A significant constraint of the findings is the dependence on a linear projection methodology, which fails to integrate scenario-based variations or uncertainty ranges, consequently diminishing the reliability of long-term predictions.

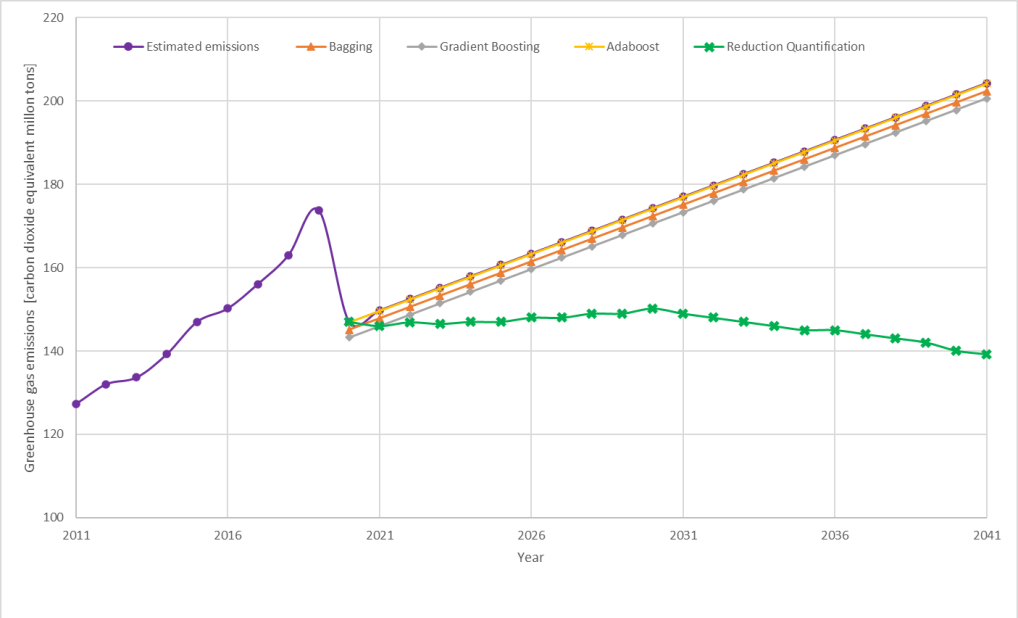


Figure 15. GHG predictions and impact quantification up to year 2041.

It has been observed that the emissions are expected to rise by 10% in every five years. Thus, in the time span of 20 years from 2021, the emissions will reach up to 205 units from 150 units, if measured in CO_2 equivalent million tons. All three regression models predicted a GHG emission rise of around 35% in 22-year period, which is very alarming and needs to be taken care of. If the necessary mitigation strategy is not implemented on time, the environmental crisis will rise to a very severe situation in Bangladesh after 22 years.

Based on the results achieved, a reduction quantification approach has been implemented, which has also been presented in Figure 15. It has been noticed that the GHG emission will remain the same and even drop below the current quantity if the actions are taken timely and appropriately. Bangladesh's Intended Nationally Determined Contributions (INDC) provide a detailed plan for reducing greenhouse gas (GHG) emissions, promoting the shift to a low-carbon, climate-resilient economy, and aiming to achieve middle-income status [67]. Bangladesh has set ambitious targets for its power generation, aiming to achieve 40% clean energy coverage, including 10% from renewables, by 2041. In accordance with its revised National Determined Contribution (NDC) as of August 2021, the nation aims to attain a renewable capacity of 4.1 GW by 2030, with a specific emphasis

on generating around 2.3 GW from solar energy. The estimated GHG emission in year 2030 is 174.28 CO_2 equivalent million tons. The INDC targets to limit emissions to 150.22 CO_2 equivalent million tons. The absolute difference of CO_2 emission is 24.02 million tons. In addition, Bangladesh plans to establish a maximum of 7 gigawatts (GW) of nuclear power by 2041 [68]. The building of the first two reactors, each with a capacity of 1.2 GW, at the Rooppur nuclear facility, is presently underway.

4.6. GHG Mitigation Policies for Bangladesh

In addition to the current mitigation measures, a variety of new ones will need to be implemented in order to lower the rate of greenhouse gas emissions. To achieve its net-zero emissions targets, Bangladesh needs to develop a comprehensive plan to cut carbon emissions in all of its main economic sectors, such as the energy, transportation, agriculture and industrial sectors. Some of the strategies that are suitable for Bangladesh to reduce the carbon footprint are as follows:

- In order to manage emissions and local air pollution, fulfill its growing energy demand, and sustain economic progress, Bangladesh should diversify its energy industry using sustainable and secure resources.
- Improved nitrogen management in Bangladesh's agricultural sector can account for 60–65% of the sector's total mitigation potential, which can lower the nation's carbon emissions while boosting production efficiency.
- Bangladesh may lower its greenhouse gas emissions and accelerate economic growth by enacting laws that mitigate emissions and air pollution.
- Reducing greenhouse gas emissions can be achieved by making investments in the sustainable energy sector and increasing the use of energy-efficient technologies and clean, renewable energy sources.
- Implementing energy-efficient and circular economy solutions in the ready-made garments industry can help reduce GHG emissions.
- The natural resource base of Bangladesh is severely degrading due to deforestation and biodiversity loss, which affects people's ability to make a living. Reducing carbon stores in vegetation and soils and strengthening the conservation of forest resources can help reduce greenhouse gas emissions.
- More sophisticated ML models can provide policymakers with accurate predictions of future greenhouse gas emissions based on various factors. This information enables policymakers to make data-driven decisions when formulating and adjusting emission reduction strategies and policies in Bangladesh.
- ML models can highlight the most influential factors contributing to GHG emissions in Bangladesh. By understanding which variables have the greatest impact, policymakers can prioritize interventions in sectors such as energy, population, urbanization, and economic activities to maximize emission reduction efforts.
- Different scenarios can be simulated using ML models in Bangladesh, which can help assess the potential impact of various policy measures on future emissions. This allows policymakers to explore different strategies and identify the most effective combination of interventions to achieve emission reduction targets.

5. Conclusions

This study presents ensemble machine learning models that can efficiently forecast and measure the effects of greenhouse gas emissions. Additionally, these models can identify the policy implications in the face of significant environmental transformations. The models have utilized time-series emissions data from 1970 to 2019 as well as several economic indicators, including energy consumption, population, urbanization, gross do-

mestic products, foreign direct investment, and per capita income. Three models, bagging, gradient boost, and Adaboost, have been developed, and their hyperparameters have been optimized using MPSO algorithms. The models have predicted GHG emissions up to the year 2041 efficiently, as evidenced by several statistical measurements, including R^2 score, MSE, MAE, and MAPE. Most of the models have shown R^2 score very close to 0.9, while the bagging model has outperformed others in terms of MSE, MAE, and MAPE. Energy consumption has shown the highest influence on GHG predictions for all models, with feature importance values of 0.75, 0.60, and 0.203 for gradient boost, Adaboost, and bagging models, respectively. The integration of clean energy has shown significant reductions in GHG emissions as per the nationally determined contributions of Bangladesh. This study can be an excellent guide for decision-makers to understand the complex nature of GHG emissions and achieve net-zero emissions. The limitations of the study include scenario-based projections, uncertainty simulations, and the use of only a single 80/20 data split for model validation instead of more advanced techniques such as k-fold cross-validation or bootstrapping, which should be rectified in future research to enhance the reliability and policy relevance of long-term forecasts. Furthermore, a systematic comparison and verification of the MPSO algorithm against other hyperparameter optimization techniques (such as Bayesian Optimization or Grid Search) is required in future studies to fully quantify the advantage of the selected tuning mechanism in terms of convergence speed and stability. Also, future studies could explore the dynamic interplay between evolving technological advancements and their influence on the effectiveness of clean development mechanisms in mitigating greenhouse gas emissions, within the context of changing environmental and socioeconomic landscapes, using advanced machine learning models such as deep neural networks, hybrid models, and LSTM.

Author Contributions: Conceptualization, M.S.A.; methodology, M.S.A.; software, M.S.A. and M.A.A.; validation, M.S.A., W.M.H. and M.S.; formal analysis, M.S.A. and M.A.; investigation, M.S.A.; resources, M.A.H.; data curation, M.S.A. and M.A.; writing—original draft preparation, M.S.A.; writing—review and editing, M.S.A., M.S.S., M.A.A., W.M.H., M.A., M.S. and M.A.H.; visualization, M.S.A. and W.M.H.; supervision, M.S.A. and M.A.H.; project administration, M.S.A.; funding acquisition, M.A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are available on request from the corresponding author.

Acknowledgments: This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU253452].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, J.; Dumitrescu, C.E. Flame development analysis in a diesel optical engine converted to spark ignition natural gas operation. *Appl. Energy* **2018**, *230*, 1205–1217. [CrossRef]
2. Liu, J.; Dumitrescu, C.E. 3D CFD simulation of a CI engine converted to SI natural gas operation using the G-equation. *Fuel* **2018**, *232*, 833–844. [CrossRef]
3. Mohan, R.R. Time series GHG emission estimates for residential, commercial, agriculture and fisheries sectors in India. *Atmospheric Environ.* **2018**, *178*, 73–79. [CrossRef]
4. Trends in Global CO_2 and Total Greenhouse Gas Emissions. 2021 Summary Report | PBL Netherlands Environmental Assessment Agency. Available online: https://www.pbl.nl/en/publications/trends-in-global-co2-and-total-greenhouse-gas-emissions-2021-summary-report (accessed on 2 February 2024).
5. Grigorescu, A.; Lincaru, C.; Pirciog, C.S. The Impact of Cloud Computing on Mass and Energy Flows: Greenhouse Gas Emissions in the IT and Communications Sectors at the European Level (2014–2021). *Processes* **2025**, *13*, 1808. [CrossRef]

6. Nunes, L.J.R. The Rising Threat of Atmospheric CO_2: A Review on the Causes, Impacts, and Mitigation Strategies. *Environments* **2023**, *10*, 66. [CrossRef]
7. Orkomi, A.A. Historical trends, underlying factors and the 2035 horizon situation of GHG emission in 16 Middle Eastern nations. *Nations Energy Sustain. Dev.* **2025**, *86*, 101693. [CrossRef]
8. Alam, M.S.; Al-Ismail, F.S.; Salem, A.; Abido, M.A. High-level penetration of renewable energy sources into grid utility: Challenges and Solutions. *IEEE Access* **2020**, *8*, 190277–190299. [CrossRef]
9. Raimi, D. The Greenhouse Gas Impacts of Increased US Oil and Gas Production. 2019. Available online: https://www.rff.org/publications/working-papers/greenhouse-gas-impacts-increased-us-oil-and-gas-production/ (accessed on 30 October 2025).
10. Bayomi, N.; Fernandez, J.E. Towards Sustainable Energy Trends in the Middle East: A Study of Four Major Emitters. *Energies* **2019**, *12*, 1615. [CrossRef]
11. Raihan, A.; Muhtasim, D.A.; Farhana, S.; Hasan, A.U.; Pavel, M.I.; Faruk, O.; Rahman, M.; Mahmood, A. An econometric analysis of Greenhouse gas emissions from different agricultural factors in Bangladesh. *Energy Nexus* **2023**, *9*, 100179. [CrossRef]
12. Raihan, A.; Muhtasim, D.A.; Farhana, S.; Hasan, A.U.; Pavel, M.I.; Faruk, O.; Rahman, M.; Mahmood, A. Nexus between economic growth, energy use, urbanization, agricultural productivity, and carbon dioxide emissions: New insights from Bangladesh. *Energy Nexus* **2022**, *8*, 100144. [CrossRef]
13. Patra, A.K. Prediction of enteric methane emission from cattle using linear and non-linear statistical models in tropical production systems. *Mitig. Adapt. Strat. Glob. Change* **2017**, *22*, 629–650. [CrossRef]
14. Biswas, J.C.; Haque, M.M.; Hossain, M.B.; Maniruzzaman, M.; Zahan, T.; Rahman, M.M.; Sen, R.; Ishtiaque, S.; Chaki, A.K.; Ahmed, I.M.; et al. Seasonal Variations in Grain Yield, Greenhouse Gas Emissions and Carbon Sequestration for Maize Cultivation in Bangladesh. *Sustainability* **2022**, *14*, 9144. [CrossRef]
15. Islam, M.M.; Kundu, G.K.; Khan, M.I. Aquaculture: A Faster Growing Greenhouse Gas Emission Sector in Bangladesh. *Int. J. Ecol. Dev.* **2020**, *35*, 44–57.
16. Debnath, K.B.; Mourshed, M. Why is Bangladesh's electricity generation heading towards a GHG emissions-intensive future? *Carbon Manag.* **2022**, *13*, 216–237. [CrossRef]
17. Rashid, N.; Kabir, H. Greenhouse Gas Emission Reduction through Electricity Generation from Solar Photovoltaic Systems: A Study in Dhaka. *Dhaka Univ. J. Earth Environ. Sci.* **2024**, *12*, 1–8. [CrossRef]
18. Habib, M.A.; Islam, S.M.M.; Haque, A.; Hassan, L.; Ali, Z.; Nayak, S.; Dar, M.H.; Gaihre, Y.K. Effects of Irrigation Regimes and Rice Varieties on Methane Emissions and Yield of Dry Season Rice in Bangladesh. *Soil Syst.* **2023**, *7*, 41. [CrossRef]
19. Liu, F.; Shen, X.; Gao, X.; Zhang, F.; Luo, X.; Liu, Y.; Yang, Y.; Yang, W.; Liang, T.; Wang, C.; et al. An innovative integrated management strategy drives sustainable vegetable production in southwest China: Higher yield with reduced net GHG emissions. *Eur. J. Agron.* **2025**, *169*, 127703. [CrossRef]
20. van Beek, L.; Hajer, M.; Pelzer, P.; van Vuuren, D.; Cassen, C. Anticipating futures through models: The rise of Integrated Assessment Modelling in the climate science-policy interface since 1970. *Glob. Environ. Change* **2020**, *65*, 102191. [CrossRef]
21. Kotlar, A.M.; Singh, J.; Kumar, S. Prediction of greenhouse gas emissions from agricultural fields with and without cover crops. *Soil Sci. Soc. Am. J.* **2022**, *86*, 1227–1240. [CrossRef]
22. Kalra, S.; Lamba, R.; Sharma, M. Machine learning based analysis for relation between global temperature and concentrations of greenhouse gases. *J. Inf. Optim. Sci.* **2020**, *41*, 73–84. [CrossRef]
23. Al-Omar, M.K.; Hameed, M.M.; Al-Ansiri, N.; Razali, S.F.M.; Al-Saadi, M.A. Short-, Medium-, and Long-Term Prediction of Carbon Dioxide Emissions using Wavelet-Enhanced Extreme Learning Machine. *Civ. Eng. J.* **2023**, *9*, 815–834. [CrossRef]
24. Sharma, S.; Saxena, A.K.; Bansal, M. Forecasting of GHG (Greenhouse Gas) Emission Using (ARIMA) Data Driven Intelligent Time Series Predicting Approach. In Proceedings of the 7th International Conference on Communication and Electronics Systems, ICCES 2022, Coimbatore, Tamil Nadu, India, 22–24 June 2022; pp. 315–322. [CrossRef]
25. Magazzino, C.; Mele, M.; Schneider, N. A machine learning approach on the relationship among solar and wind energy production, coal consumption, GDP, and CO_2 emissions. *Renew. Energy* **2021**, *167*, 99–115. [CrossRef]
26. 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]
27. Atasoy, D. A Study on Decision Tree Optimization Algorithm. Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi. Available online: https://www.academia.edu/108039688/A_Study_on_Decision_Tree_Optimization_Algorithm (accessed on 2 February 2024).
28. Ahmed, Y.; Maya, A.A.S.; Akhtar, P.; Alam, S.; AlMohamadi, H.; Islam, N.; Alharbi, O.A.; Rahman, S.M. A novel interpretable machine learning and metaheuristic-based protocol to predict and optimize ciprofloxacin antibiotic adsorption with nano-adsorbent. *J. Environ. Manag.* **2024**, *370*, 122614. [CrossRef]
29. Decision Tree Algorithm in Machine Learning—Javatpoint. Available online: https://www.tpointtech.com/machine-learning-decision-tree-classification-algorithm (accessed on 2 February 2024).

30. Aziz, S.; Chowdhury, S.A. Analysis of agricultural greenhouse gas emissions using the STIRPAT model: A case study of Bangladesh. *Environ. Dev. Sustain.* **2023**, *25*, 3945–3965. [CrossRef]
31. Chowdhury, S.; Rubi, M.A.; Bijoy, M.H.I. Application of Artificial Neural Network for Predicting Agricultural Methane and CO2 Emissions in Bangladesh. In Proceedings of the 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021, Khargpur, India, 6–8 July 2021. [CrossRef]
32. Das, N.G.; Sarker, N.R.; Haque, M.N. An estimation of greenhouse gas emission from livestock in Bangladesh. *J. Adv. Vet. Anim. Res.* **2020**, *7*, 133–140. [CrossRef]
33. Khatun, M.; Salma, U.; Sultana, I.; Hasan, M.M. Carbon Dioxide Emission from Fossil Fuel: A Procedure for Building a Predictive Model. In Proceedings of the 3rd International Conference on Electrical, Computer and Telecommunication Engineering, ICECTE 2019, Rajshahi, Bangladesh, 26–28 December 2019; pp. 177–180. [CrossRef]
34. Rahman, M.M.; Shafiullah, M.; Alam, M.S.; Rahman, M.S.; Alsanad, M.A.; Islam, M.M.; Islam, M.K.; Rahman, S.M. Decision Tree-Based Ensemble Model for Predicting National Greenhouse Gas Emissions in Saudi Arabia. *Appl. Sci.* **2023**, *13*, 3832. [CrossRef]
35. Al-Ismail, F.S.; Alam, M.S.; Shafiullah, M.; Hossain, M.I.; Rahman, S.M. Impacts of Renewable Energy Generation on Greenhouse Gas Emissions in Saudi Arabia: A Comprehensive Review. *Sustainability* **2023**, *15*, 5069. [CrossRef]
36. Chen, Z.; Guo, Y.; Guo, C. Prediction of GHG emissions from Chengdu Metro in the construction stage based on WOA-DELM. *Tunn. Undergr. Space Technol.* **2023**, *139*, 105235. [CrossRef]
37. Hamdan, A.; Al-Salaymeh, A.; Alhamad, I.M.; Ikemba, S.; Raphael, D.; Ewim, E. Predicting future global temperature and greenhouse gas emissions via LSTM model. *Sustain. Energy Res.* **2023**, *10*, 21. [CrossRef]
38. AlShafeey, M.; Rashdan, O. Quantifying the impact of energy consumption sources on GHG emissions in major economies: A machine learning approach. *Energy Strat. Rev.* **2023**, *49*, 101159. [CrossRef]
39. Gütschow, J.; Jeffery, M.L.; Gieseke, R.; Gebel, R.; Stevens, D.; Krapp, M.; Rocha, M. The PRIMAP-hist national historical emissions time series. *Earth Syst. Sci. Data* **2016**, *8*, 571–603. [CrossRef]
40. Bangladesh | Data. Available online: https://data.worldbank.org/country/BD (accessed on 6 February 2024).
41. Energy Statistics Data Browser—Data Tools—IEA. Available online: https://www.iea.org/data-and-statistics/data-tools/energy-statistics-data-browser?country=WORLD&fuel=Energy%20supply&indicator=TESbySource (accessed on 8 October 2024).
42. Country Profile—amaWebClient. Available online: https://unstats.un.org/unsd/snaama/countryprofile (accessed on 8 October 2024).
43. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140.
44. González, S.; García, S.; Del Ser, J.; Rokach, L.; Herrera, F. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Inf. Fusion* **2020**, *64*, 205–237. [CrossRef]
45. Alam, M.S.; Tiwari, S.P.; Rahman, S.M. Optimized Ensemble Machine Learning Models for Predicting Phytoplankton Absorption Coefficients. *IEEE Access* **2024**, *12*, 5760–5769. [CrossRef]
46. Min, H.; Luo, X. Calibration of soft sensor by using Just-in-time modeling and AdaBoost learning method. *Chin. J. Chem. Eng.* **2016**, *24*, 1038–1046.
47. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
48. Alam, M.S.; Al-Ismail, F.S.; Hossain, M.S.; Rahman, S.M. Ensemble Machine-Learning Models for Accurate Prediction of Solar Irradiation in Bangladesh. *Processes* **2023**, *11*, 908. [CrossRef]
49. Fang, J.; Liu, W.; Chen, L.; Lauria, S.; Miron, A.; Liu, X. A survey of algorithms, applications and trends for particle swarm optimization. *Int. J. Netw. Dyn. Intell.* **2023**, *2023*, 24–50. [CrossRef]
50. Alam, M.S.; Al-Ismail, F.S.; Al-Sulaiman, F.A.; Abido, M.A. Energy management in DC microgrid with an efficient voltage compensation mechanism. *Electr. Power Syst. Res.* **2022**, *214*, 108842. [CrossRef]
51. Eberhart, R.C.; Shi, Y. Comparing Inertia Weights and Constriction Factors in Particle Swarm Optimization. In Proceedings of the 2000 Congress on Evolutionary Computation, CEC00 (Cat. No. 00TH8512), San Diego, CA, USA, 16–19 July 2000; pp. 84–88.
52. Al-Awami, A.T.; Zerguine, A.; Cheded, L.; Zidouri, A.; Saif, W. A new modified particle swarm optimization algorithm for adaptive equalization. *Digit. Signal Process.* **2011**, *21*, 195–207. [CrossRef]
53. Alam, M.S.; Al-Ismail, F.S.; Abido, M.A. PV/Wind-Integrated Low-Inertia System Frequency Control: PSO-Optimized Fractional-Order PI-Based SMES Approach. *Sustainability* **2021**, *13*, 7622. [CrossRef]
54. Darwish, H.; Al Hmoud, I.W.; Turlapaty, A.C.; Gokaraju, B. Predicting the future climate: Integrating renewable energy and machine learning to address temperature and GHG emissions. *Energy Rep.* **2025**, *14*, 2399–2419. [CrossRef]
55. Zhou, Y.; Zheng, S.; Liu, Z.; Wen, T.; Ding, Z.; Yan, J.; Zhang, G. Passive and active phase change materials integrated building energy systems with advanced machine-learning based climate-adaptive designs, intelligent operations, uncertainty-based analysis and optimisations: A state-of-the-art review. *Renew. Sustain. Energy Rev.* **2020**, *130*, 109889. [CrossRef]

56. Ahmed, Y.; Rahman, M.; Alam, S.; Miah, M.I.; Choudhury, S.H.; Alharbi, O.A.; Akhtar, P.; Rahman, S.M. Harnessing neural network model with optimization for enhanced ciprofloxacin antibiotic adsorption from contaminated water: A transparent and objective framework. *J. Water Process. Eng.* **2024**, *65*, 105724. [CrossRef]
57. Heidari, A.; Davtalab, J.; Sargazi, M.A.; Piri, J. Machine Learning-Enhanced Assessment of Natural versus Artificial Shade Cooling Performance in Hot-Arid Climates: SVR Modeling with Hybrid GA-PSO Optimization. *Build. Environ.* **2025**, *287*, 113789. [CrossRef]
58. Islam, M.S.S.; Ghosh, P.; Faruque, M.O.; Islam, M.R.; Hossain, M.A.; Alam, M.S.; Islam Sheikh, M.R. Optimizing Short-Term Photovoltaic Power Forecasting: A Novel Approach with Gaussian Process Regression and Bayesian Hyperparameter Tuning. *Processes* **2024**, *12*, 546. [CrossRef]
59. Choi, W.; Yoo, E.; Seol, E.; Kim, M.; Song, H.H. Greenhouse gas emissions of conventional and alternative vehicles: Predictions based on energy policy analysis in South Korea. *Appl. Energy* **2020**, *265*, 114754. [CrossRef]
60. Chu, B.; Duncan, S.; Papachristodoulou, A.; Hepburn, C. Analysis and control design of sustainable policies for greenhouse gas emissions. *Appl. Therm. Eng.* **2013**, *53*, 420–431. [CrossRef]
61. Tudor, C.; Sova, R.; Gegov, A.; Jafari, R. Benchmarking GHG Emissions Forecasting Models for Global Climate Policy. *Electronics* **2021**, *10*, 3149. [CrossRef]
62. Shams, S.; Sahu, J.N.; Rahman, S.M.S.; Ahsan, A. Sustainable waste management policy in Bangladesh for reduction of greenhouse gases. *Sustain. Cities Soc.* **2017**, *33*, 18–26. [CrossRef]
63. Alam, S.; Hossain, A.; Shafiullah, D.; Islam, A.; Choudhury, M.; Faruque, O.; Abido, M.A. Renewable energy integration with DC microgrids: Challenges and opportunities. *Electr. Power Syst. Res.* **2024**, *234*, 110548. [CrossRef]
64. Nazar, S.; Yang, J.; Wang, X.-E.; Khan, K.; Amin, M.N.; Javed, M.F.; Althoey, F.; Ali, M. Estimation of strength, rheological parameters, and impact of raw constituents of alkali-activated mortar using machine learning and SHapely Additive exPlanations (SHAP). *Constr. Build. Mater.* **2023**, *377*, 131014. [CrossRef]
65. Interpretable Machine Learning. Available online: https://christophm.github.io/interpretable-ml-book/ (accessed on 21 October 2024).
66. Alam, M.S.; Rahman, S.M.; Sattar, K.; Islam, M.N.; Choudhury, M.S.H.; Assi, K.J.; Al-Ratrout, N.T. Traffic Crash-Severity Prediction using Interpretable Optimized Ensemble Models. *Int. J. Intell. Transp. Syst. Res.* **2025**, *2025*, 1–23. [CrossRef]
67. Intended Nationally Determined Contributions (INDC) 2015. National Designated Authority to GCF. Available online: https://unfccc.int/sites/default/files/NDC/2022-06/INDC_2015_of_Bangladesh.pdf (accessed on 6 February 2024).
68. Bangladesh Targets 40% of Clean Power Generation by 2041 | Enerdata. Available online: https://www.enerdata.net/publications/daily-energy-news/bangladesh-targets-40-clean-power-generation-2041.html (accessed on 6 February 2024).

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.