

Article

# Predicting CO<sub>2</sub> Emissions with Advanced Deep Learning Models and a Hybrid Greylag Goose Optimization Algorithm

Amel Ali Alhussan <sup>1</sup>, Marwa Metwally <sup>2,3</sup> and S. K. Towfek <sup>3,4,\*</sup>

<sup>1</sup> Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

<sup>2</sup> Jadara University Research Center, Jadara University, Irbid 21110, Jordan

<sup>3</sup> Computer Science and Intelligent Systems Research Center, Blacksburg, VA 24060, USA

<sup>4</sup> Applied Science Research Center, Applied Science Private University, Amman, Jordan

\* Correspondence: sktowfek@jcsis.org

**Abstract:** Global carbon dioxide (CO<sub>2</sub>) emissions are increasing and present substantial environmental sustainability challenges, requiring the development of accurate predictive models. Due to the non-linear and temporal nature of emissions data, traditional machine learning methods—which work well when data are structured—struggle to provide effective predictions. In this paper, we propose a general framework that combines advanced deep learning models (such as GRU, Bidirectional GRU (BIGRU), Stacked GRU, and Attention-based BIGRU) with a novel hybridized optimization algorithm, GGBERO, which is a combination of Greylag Goose Optimization (GGO) and Al-Biruni Earth Radius (BER). First, experiments showed that ensemble machine learning models such as CatBoost and Gradient Boosting addressed static features effectively, while time-dependent patterns proved more challenging to predict. Transitioning to recurrent neural network architectures, mainly BIGRU, enabled the modeling of sequential dependence on emissions data. The empirical results show that the GGBERO-optimized BIGRU model produced a Mean Squared Error (MSE) of  $1.0 \times 10^{-5}$ , the best tested approach. Statistical methods like the Wilcoxon Signed Rank Test and ANOVA were employed to validate the framework's effectiveness in improving the evaluation, confirming the significance and robustness of the improvements due to the framework. In addition to improving the accuracy of CO<sub>2</sub> emissions forecasting, this integrated approach delivers interpretable explanations of the significant factors of CO<sub>2</sub> emissions, aiding policymakers and researchers focused on climate change mitigation in data-driven decision-making.



Academic Editors: Lei He, Xiaolu Liu, Lining Xing and Guohua Wu

Received: 8 April 2025

Revised: 27 April 2025

Accepted: 28 April 2025

Published: 30 April 2025

**Citation:** Alhussan, A.A.; Metwally, M.; Towfek, S.K. Predicting CO<sub>2</sub> Emissions with Advanced Deep Learning Models and a Hybrid Greylag Goose Optimization Algorithm. *Mathematics* **2025**, *13*, 1481. <https://doi.org/10.3390/math13091481>

**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/>).

**Keywords:** CO<sub>2</sub> emissions; deep learning; optimization algorithms; feature selection; recurrent neural networks; gated recurrent unit (GRU)

**MSC:** 68T01; 68T07; 68T20

## 1. Introduction

Carbon dioxide (CO<sub>2</sub>) emissions worldwide continue to rise, and as a result, environmental sustainability is at risk; using correct and accurate models to predict emission rates of CO<sub>2</sub> is central to reducing the impact of climate change. Sophisticated forecasting models are essential not only for the development of specific strategies but also for the tracking of emission dynamics on a macro level [1,2]. The traditional way of developing prediction models is insufficient for such a case in general because of the nature of environmental data, which carries non-linear and random characteristics and is time-variable.

These limitations have led to a search for better methods, especially those reinforced with artificial intelligence (AI) techniques based on fundamental concepts of machine learning (ML), which have been improved with optimization techniques [3].

Machine learning models have recently been used widely in environmental data analysis because of their ability to find patterns and correlations in the datasets. However, the performance of these models has a strong link to the algorithms used in them, mainly when it comes to time-variant sequences like CO<sub>2</sub> emission [4–6]. This research's first approaches apply a set of preliminary yet sufficient ML models, namely Cat Boost, Gradient Boosting, Extra Trees, and XGBoost, which are well known for their stability and versatility in working with structured data. These ensemble learning methods, developed based on gradient-based optimization, offer an excellent baseline for the predictions made. However, they cannot model long-term dependencies effectively as they are inadequate in considering essential sequential details of the time series dataset [7–9].

To address these limitations, this study advances to sophisticated Recurrent Neural Network (RNN) architectures, particularly Gated Recurrent Unit (GRU) variants. Three advanced architectures are emphasized: Bidirectional GRU (BIGRU), Stacked GRU, and Attention-enhanced BIGRU. BIGRU processes sequences bidirectionally, enhancing temporal relation capture [10], while stacked layers enable hierarchical feature learning. Attention mechanisms further improve focus on critical time steps [11–13].

The study introduces two novel optimization algorithms: Greylag Goose Optimization (GGO) and Al-Biruni Earth Radius (BER). GGO mimics avian collective behavior to balance exploration–exploitation [14–16], while BER employs geometric principles for convergence in noisy optimization spaces [17–19]. Their hybridization in GGBERO enables robust feature selection and model tuning.

The effectiveness of the proposed models and the optimization strategies is assessed using statistical techniques such as the Wilcoxon Signed Rank Test and the Analysis of Variance (ANOVA). The Wilcoxon test is useful when rating pairs of samples, allowing the treatment of the interspersed improvements between the models to be significantly improved [20–22]. In contrast, ANOVA makes it easier to compare several models based on the given criteria and provide a general evaluation of the difference in performance. Such statistical comparisons are essential to ensure that all the improvements recorded are not because of random chances but the superior methods used in this research [23–25].

This study proposes a unified framework for CO<sub>2</sub> emission prediction using advanced machine learning structures and optimization techniques. The enhanced precision and reliability of the BIGRU model, mainly when the GGBERO algorithm is applied, is a significant finding. The study's contributions to machine learning and environmental science are substantial, demonstrating the potential of AI-derived models in addressing global issues.

The subsequent sections of this research provide a comprehensive literature review of earlier studies, details of the dataset used, selection of the features, GRU models, and optimization techniques. The model's success is critically evaluated in the experimental results section, and the key findings are discussed, along with suggestions for further studies.

## 2. Literature Review

In the literature review, the author focuses on comparing various research works that centered on forecasting carbon emissions with the various models and methodologies applied to enhance the predictive abilities of the forecast models. When governments across the globe press on the need to set ever-higher carbon reduction targets, accurate emission forecasts are inevitable. Of late, with the help of new-age techniques in machine learning and deep learning, efforts have been made to forecast better results. The following

section reviews and compares these models across various sectors and regions, and how deep learning may provide better acknowledgment in future emission prediction. Also, the authors present studies that address carbon emissions in transportation, construction, and industrial segments of the economy while stressing the use of model performance indicators such as RMSE, MAPE, and  $R^2$ . This review focuses on setting out an accurate picture of the role these models can play in achieving improvements in carbon emission mitigation methods.

Authors of [26] compared the effectiveness of econometric, machine learning, and deep learning models for carbon emission forecasting. Their findings indicate that heuristic neural networks (a deep learning approach) demonstrate higher predictability for future emission forecasting compared to econometric models, though econometric models are more suitable for estimating changes due to specific factors. Also paper [27] focused on the building and construction sector in China, analyzing emissions across 30 provinces with nine machine learning regression models. Their results show that a stacking ensemble regression model outperformed others, identifying urbanization and population as key drivers of emissions, thus supporting the development of targeted low-carbon policies.

Authors of [28] addressed embodied carbon emissions in construction by applying artificial neural networks, support vector regression, and extreme gradient boosting to estimate emissions at the design phase. Their models, tested on 70 projects, achieved strong interpretability ( $R^2 > 0.7$ ) and low error, supporting practical tools for emission estimation and reduction during construction. In addition, authors of [29] developed an interpretable multi-stage forecasting framework using SHAP to analyze energy consumption and CO<sub>2</sub> emissions in the UK transport sector. Their results indicate that road carbon intensity is the most significant predictor, while population and GDP per capita have less impact than previously thought.

Authors of [30] utilized XGBoost to analyze real-world driving data for heavy-duty vehicles in the EU, demonstrating that on-board monitoring data enables more accurate CO<sub>2</sub> emission predictions than traditional fuel-based methods. Also paper [31] introduced an interpretable machine learning approach using land use data to predict emissions in the Yangtze River Delta. Their Extra Tree Regression Optimization model achieved high accuracy ( $R^2 = 0.99$  on training, 0.86 on test data) and revealed spatial clusters of emissions, with industrial land use contributing to regional hotspots.

Authors of [32] evaluated several machine learning models—including linear regression, ARIMA, and shallow and deep neural networks—for long-term CO<sub>2</sub> emissions forecasting in the building sector across multiple countries. Deep neural networks provided the best long-term prediction performance. Paper [33] proposed a hybrid deep learning framework combining gated recurrent units (GRUs) and graph convolutional networks (GCNs) to capture both temporal and spatial dependencies in Chinese urban clusters. Their model outperformed baselines in both single- and multi-step forecasts and demonstrated strong generalizability.

Authors of [34] developed and compared nine machine learning regression models for national-level CO<sub>2</sub> emissions, finding that optimized Gaussian Process Regression achieved the highest accuracy ( $R^2 = 0.9998$ ). Paper [35] used ARIMA, SARIMAX, Holt-Winters, and LSTM models to predict India's CO<sub>2</sub> emissions, with LSTM achieving the lowest MAPE (3.101%) and RMSE (60.635), confirming its suitability for emission forecasting.

Paper [36] applied deep learning, support vector machines, and artificial neural networks to forecast transportation-related CO<sub>2</sub> emissions and energy demand in Turkey, finding strong correlations between economic indicators and emissions, and predicting significant increases in both metrics over the next 40 years. Also Authors of [37] used reinforcement learning to optimize ship routes, reducing fuel consumption and emissions.

The DDPG algorithm achieved the best performance, demonstrating the potential of RL for emission reduction in shipping.

Authors of [38] forecasted greenhouse gas emissions in Turkey’s electricity sector using deep learning and ANN, achieving high accuracy across several metrics, and highlighting the rapid growth of GHG emissions in recent decades. Also paper [39] used LSTM models for high-frequency greenhouse gas emission prediction in transport networks, outperforming clustering and ARIMA models and supporting the use of deep learning for detailed, real-time emission forecasting. In addition paper [40] improved CO<sub>2</sub> emission prediction in China by combining factor analysis with a PSO-optimized extreme learning machine (PSO-ELM), achieving higher accuracy than conventional ELM and backpropagation neural networks. Their approach supports more effective economic policy design for emission reduction.

Table 1 shows a comparative analysis of various studies on forecasting carbon emissions using different models and methodologies. The table captures key aspects, including the type of model used (deep learning, machine learning, econometric models), the main sector or region studied, performance metrics, and key findings. The comparison highlights that deep learning and machine learning models often exhibit higher accuracy and better prediction capabilities, particularly when applied to spatial–temporal data or in complex systems like transportation and construction. Moreover, econometric models tend to excel at estimating changes due to specific factors but may lack the predictive power of more advanced models. These findings align with the general trend in the literature, where machine learning models, especially ensemble and deep learning approaches, are increasingly favored for their adaptability and improved performance across diverse datasets and sectors.

**Table 1.** Comparative analysis of carbon emissions forecasting models and key findings.

Study	Model Type	Sector/Region	Key Metrics
[26]	Heuristic Neural Network	Carbon Emissions Forecasting	Predictability, Emission Estimation
[27]	Stacking Ensemble Regression	Construction Sector, China	RMSE, R <sup>2</sup> , MAPE
[28]	ANN, SVR, XGBoost	Building Projects, China	R <sup>2</sup> > 0.7, Error < 5.33%
[29]	SHAP, ML Ensemble	Transport, UK	Interpretability, Accuracy
[30]	XGBoost	Road Emissions, EU	CO <sub>2</sub> Emission Estimates
[31]	Extra Tree Regression, SHAP	Yangtze River Delta, China	R <sup>2</sup> = 0.99 (Training), R <sup>2</sup> = 0.86 (Test)
[32]	ARIMA, Neural Networks	Global (Building Sector)	Multivariate Accuracy
[33]	GRU, GCN	Urban Clusters, China	Spatiotemporal Prediction Accuracy
[34]	Gaussian Process Regression	National Level (CO <sub>2</sub> Forecasting)	MSE = 106.68, RMSE = 10.328, R <sup>2</sup> = 0.9998
[35]	LSTM, SARIMAX	India	MAPE = 3.101%, RMSE = 60.635
[36]	DL, SVM, ANN	Transport, Turkey	R <sup>2</sup> > 0.86, rRMSE ≈ 10%
[37]	RL (DDPG, DQN)	Shipping Industry	Fuel Consumption, Emissions Reduction
[38]	ANN, DL	Electricity Sector, Turkey	RMSE, rRMSE, R <sup>2</sup>
[39]	LSTM, ARIMA	Transportation (Link-Level)	RMSE = 30, Coefficient Accuracy
[40]	PSO-ELM (Extreme Learning Machine with Particle Swarm Optimization)	Carbon Emissions, China (Hebei)	Prediction Accuracy, Factor Analysis

As illustrated in this literature review section, various mitigation techniques and technologies are highlighted above to show how diverse they are for reducing CO<sub>2</sub> from vehicles. The studies we analyzed above describe various strategies to reduce car emissions, such as regulation, technological, and other policy options. Though there has been some success in making transportation systems cleaner and more efficient, issues still need to be addressed, including the lack of infrastructure and customer acceptance to implement policies. A multidimensional strategy incorporating technological innovation, supportive policies, and social participation will have to be employed in the future to achieve significant

reductions of CO<sub>2</sub> emissions from vehicles collectively. The insights gathered from these studies thus create room for a greener and more effective transit environment. Policymakers can use the studies, business stakeholders, and researchers to assist them in decision-making as the world strives toward sustainability, promoting climate resilience.

Following the literature review, the key research gaps identified in this study are summarized as follows:

- Inadequate handling of non-linear, time-dependent patterns in CO<sub>2</sub> emissions data by traditional models;
- Limited capacity of existing methods to capture long-term sequential relationships;
- Insufficient integration of advanced optimization techniques for feature selection and hyperparameter tuning;
- Absence of a unified framework combining deep learning architectures with hybrid optimization strategies.

To address these identified research gaps, this paper proposes the following methodological contributions:

- Development of a unified framework integrating advanced GRU architectures with hybrid GGBERO optimization;
- Demonstration of GGBERO-optimized BIGRU's superior performance (MSE:  $1.0 \times 10^{-5}$ );
- Novel hybrid optimization strategy enhancing model robustness against local optima;
- Comprehensive validation using the Wilcoxon Signed Rank Test and ANOVA;
- Interpretable insights into emission drivers through attention mechanisms.

Collectively, these contributions fill the research gaps identified and enhance the existing carbon emission prediction by presenting a reliable, interpretable, and statistically proven modeling framework. It makes a sound foundation for future studies and applications in environmental data analysis and policymaking.

### 3. Materials and Methods

This section describes the research methods used in the study and the data used to develop the models from the data collection. Therefore, this research aims to achieve a high level of accuracy and reliability in predictions for CO<sub>2</sub> emission levels and identify the hidden links between vehicle attributes and emissions through data analysis and feature selection methods coupled with state-of-the-art optimization techniques.

#### 3.1. Dataset

The dataset adopted in this research provides clear insights into how different aspects of a vehicle affect the emission of CO<sub>2</sub>, hence creating a platform to model and accurately predict the same emissions. The dataset originates in the Canadian government's official open-data portal and is seven years long, where variables are represented in 7385 rows and 12 columns, respectively. Every row is associated with a specific vehicle entry and contains crucial variables, which either interactively or non-interactively convey their impact on the vehicle's CO<sub>2</sub> emissions. Since the data are collected over several years, identifying trends and patterns that may be typical of old and newer models will be more precise [41].

The characteristics of the dataset cover aspects of basic vehicle description, including model type, transmission system, refueling type, and fuel consumption rates, all of which are crucial to defining a vehicle's CO<sub>2</sub> emission factors. All these features are named using standard four-letter alphanumeric codes for ease of analysis. For instance, the vehicle model is categorized based on its drivetrain configuration and body structure, including options such as four-wheel drive (4WD/4 × 4), all-wheel drive (AWD), flexible-fuel vehicles (FFV),

as well as short (SWB), long (LWB), and extended wheelbases (EWB). These distinctions are critical, as they directly relate to the vehicle's performance and, consequently, its emissions.

Transmission types are similarly encoded, covering a range of systems from fully automatic (A) to automated manual (AM), continuously variable (AV), and manual transmissions (M). Additionally, the dataset records the number of gears, reflecting the variability in gear configurations that can influence fuel efficiency and emissions. The inclusion of this level of detail allows for a more nuanced analysis of how transmission technology impacts CO<sub>2</sub> output, acknowledging the complex interplay between gear ratios and driving conditions.

Fuel type is another critical variable, with categories including regular gasoline (X), premium gasoline (Z), diesel (D), ethanol (E85), and natural gas (N). Each fuel type has distinct properties affecting combustion efficiency and emission levels. For instance, diesel engines, while typically more fuel-efficient than gasoline, emit higher levels of certain pollutants, making this a key consideration in emission modeling. On the other hand, ethanol-blended fuels present a different profile due to their renewable content, highlighting the diverse factors influencing the dataset.

Fuel consumption is captured in city and highway conditions, expressed in liters per 100 km (L/100 km). A combined rating that blends 55% city driving and 45% highway driving is also provided, along with an alternative measure in miles per gallon (mpg). This dual metric approach offers a comprehensive view of a vehicle's fuel economy, accommodating metric and imperial systems and enhancing the dataset's applicability across different contexts. Accurate fuel consumption data is crucial, as it serves as a proxy for understanding how efficiently a vehicle converts fuel into energy and, by extension, its emission levels.

The primary variable of interest in the dataset is CO<sub>2</sub> emissions, which has the unit of measurement as grams per kilometer (g/km). These figures depict the emissions during the commission of urban and extra-urban driving cycles, which mimic real-world driving typical of usage. The emphasis on CO<sub>2</sub> output is essential for the present dataset as this indicator is significant in global climate change talks.

The dataset is collected from various government portals, ensuring enough detail and that the level of coverage is needed for model development. The qualities of the data are well-suited to the training and testing of machine learning models to make emission predictions based on vehicle features. Including all these features makes the dataset versatile so that many aspects of the relations between car attributes and their impact on the environment can be investigated.

### 3.2. Exploratory Data Analysis

Exploratory data analysis, or EDA, is the initial cost-effective step in the data analysis process and forms the basis of understanding the entire dataset and the initial spotting of relations, patterns, or even oddities. EDA is used to analyze the data's intrinsic structure and identify valuable features that would be used in constructing the subsequent analysis. It makes it possible to establish relationships between variables and to determine the presence of outlying points or potential data distribution patterns, which are crucial for decisions throughout the analysis. To do this, EDA provides an understanding of complex data by displaying it on a heatmap, bar chart, or histogram to combine or prepare such data for a more complicated analysis [42,43].

Figure 1 represents a heat map of the correlation of several features of the raw data feed of CO<sub>2</sub> emissions and the model before the feature selection occurred, (color: dark orange = strong positive, light = weak, dark blue = strong negative). This means that this heatmap will show the detailed interaction between the variables, where the intensity of the color is a function of the interacting variables. Looking at this heatmap, we can find out if any features are very similar or those that are not contributing much to the model, and

thus guide us on which features to keep in the model and which to discard. Sometimes, indicators might be highly correlated; this could be problematic regarding model tuning, particularly within machine learning algorithms.

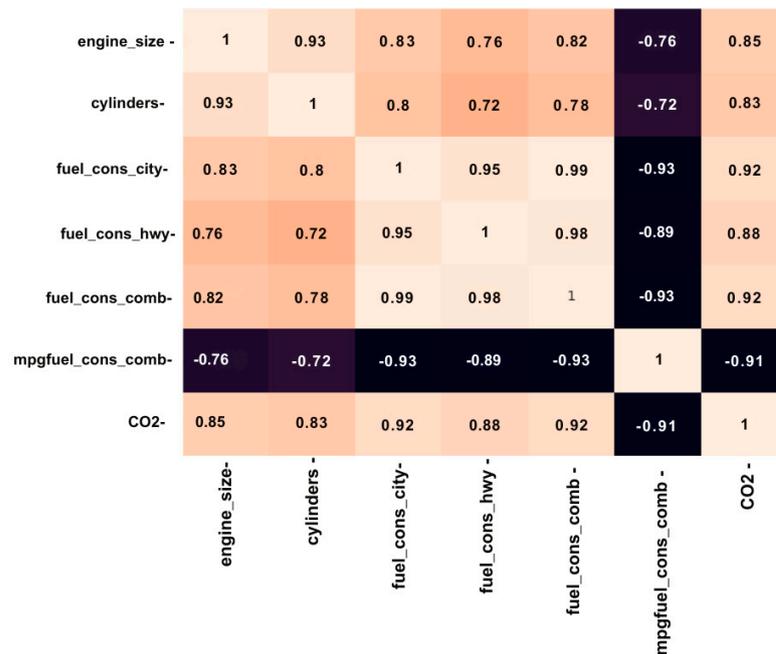


Figure 1. Heatmap of CO<sub>2</sub> dataset before feature selection.

Figure 2 presents the percent share of CO<sub>2</sub> emissions of vehicles, depending on the number of gears. It arranges vehicles according to the gear count and then evaluates this aspect’s impact on emissions. The story shows whether cars with more gears have higher or lower emissions, which helps to understand the correlation between gear layout and CO<sub>2</sub> production. This analysis is essential to understanding how vehicle mechanics, such as gear systems, influence environmental impact.

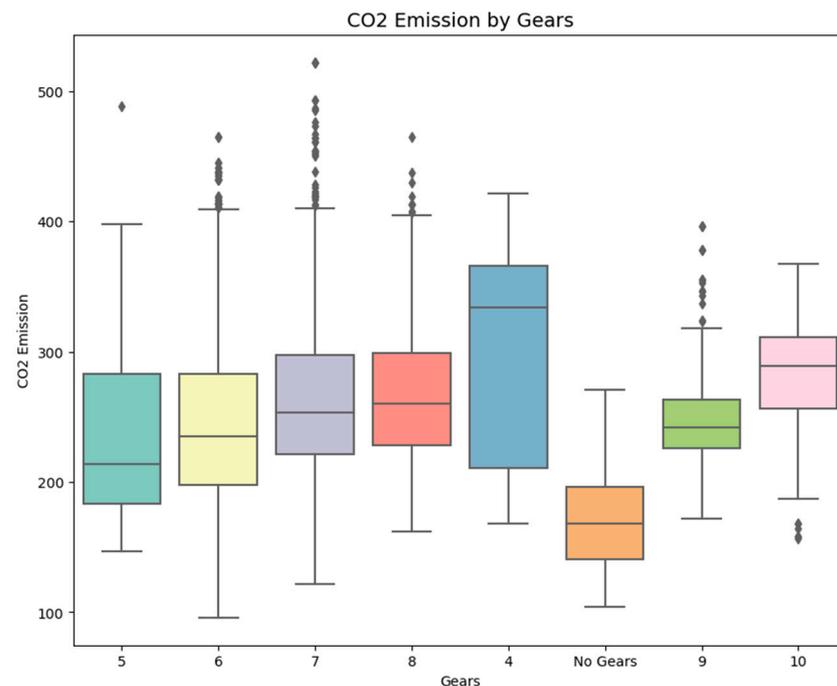


Figure 2. CO<sub>2</sub> emission by gears.

Figure 3 indicates how CO<sub>2</sub> emissions vary with each fuel type; how energy forms contribute to emission levels is equally evident. A component-wise plot of gasoline, diesel, ethanol, and natural gas compares them to give a direct perception about which fuel type is emitting high or low emissions. This visualization is beneficial for comparing the environmental impacts of various fuels and can be used to support the push for cleaner energy in automobiles.

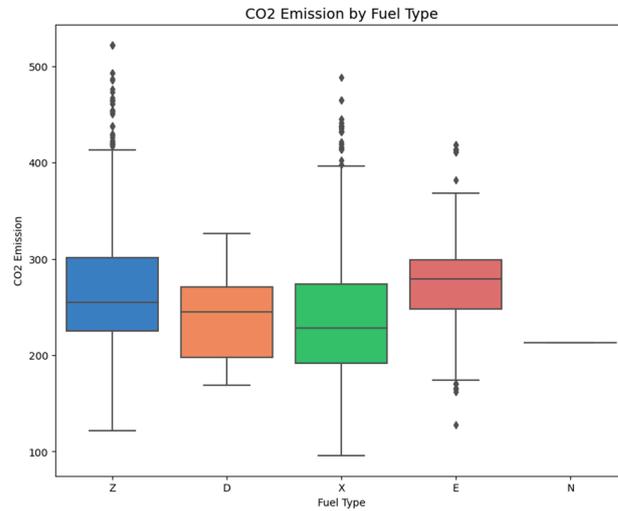


Figure 3. CO<sub>2</sub> emission by fuel type.

Figure 4 studies the trends in CO<sub>2</sub> emissions for the variants in the transmission type: automatic, manual, and continuously variable transmission (CVT). The plot enables the comparison of how transmission technology affects emissions. It is thus essential to determine these differences to evaluate which of the transmission systems is less hostile to the environment and can give some clues to the compromises between power and fuel consumption and emissions.

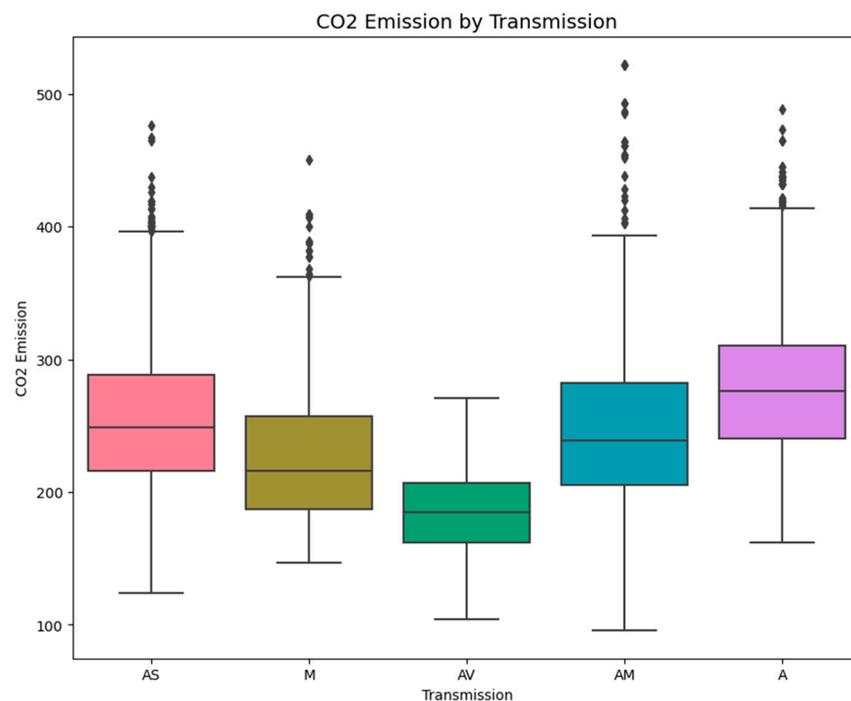
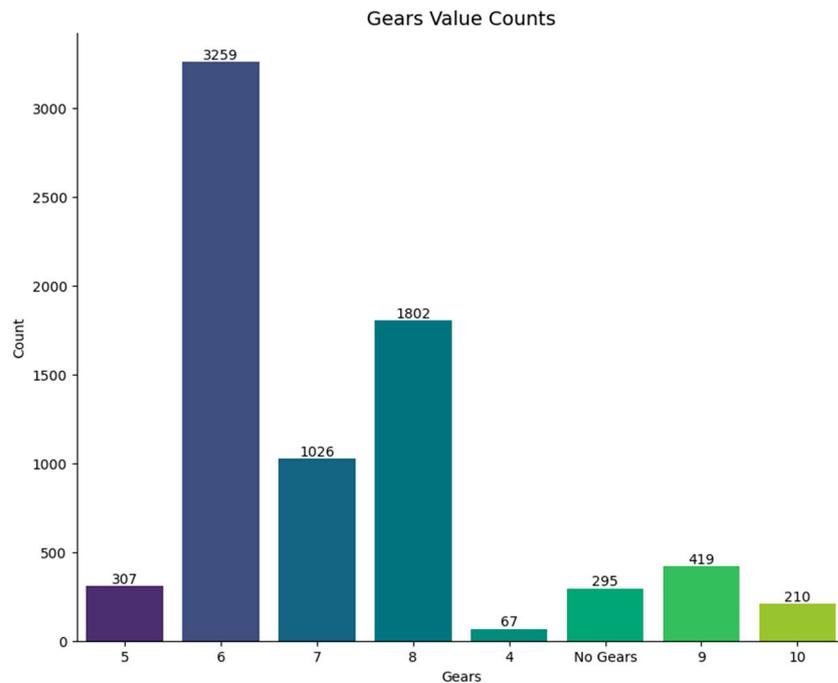


Figure 4. CO<sub>2</sub> emission by transmission.

Figure 5 presents the number of vehicles per the number of gears available in each vehicle as a frequency distribution. This bar chart also, in a simple manner, presents information on how frequently each gear configuration is encountered in the dataset. Analyzing gear counts has value because it is a way of defining tendencies in values of vehicle design and fixes for interpreting observed emissions in the analyzed dataset. It also assists in analyzing the utilization of distinct gear configurations and their effect on the environment.

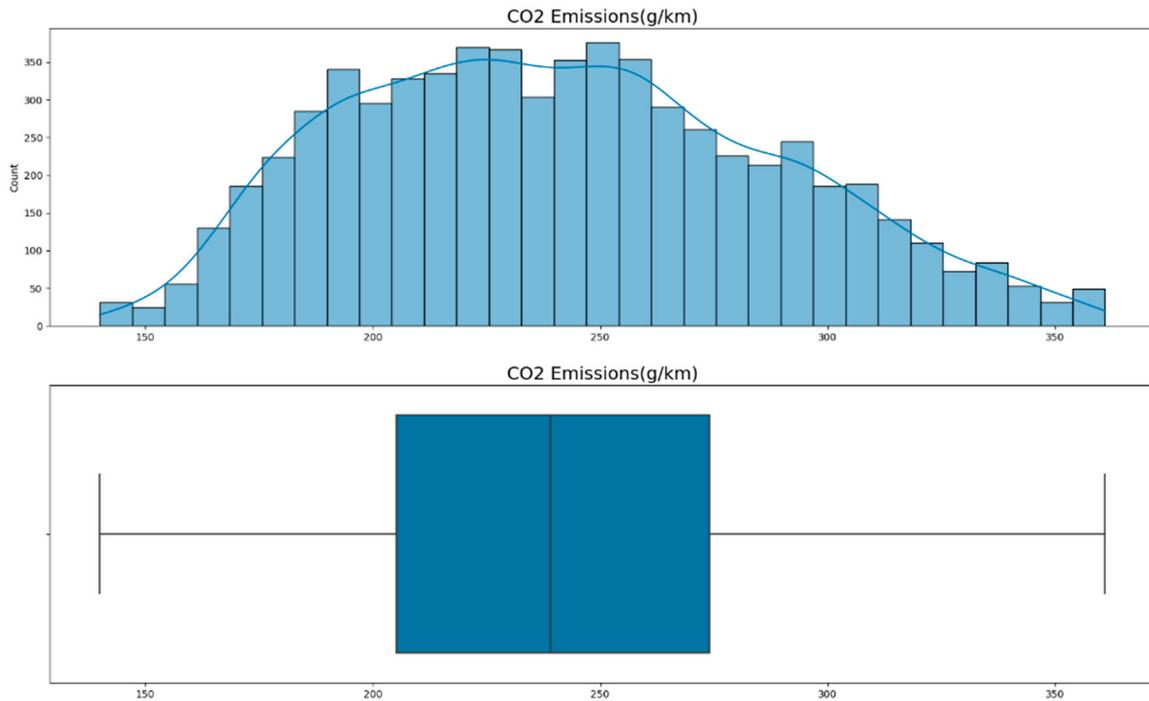


**Figure 5.** Gears value counts.

Figure 6 gives a graphical representation of the dispersion of CO<sub>2</sub> emissions in all the vehicles in the dataset. The first plot shown is a histogram with a density curve that maps out the distribution of emissions in terms of g/km and how frequent and dispersed they are. The histogram is positively skewed, suggesting that most vehicles produced between 180 and 260 g/km, and few vehicles are in the other bracket. The density curve extends the concept of the histogram more than it enlarges its value, as it smoothens the distribution and looks at the density of data in pollution concentration within any given dataset while amplifying any underlying trends.

The boxplot gives a brief description of the distribution's most essential characteristics. It provides a median CO<sub>2</sub> emission value indicated by the line inside the box and the interquartile range (IQR), which is the box's overall spread. The whiskers go up to the minimum and maximum of 1. Meanwhile, 'Outliers' are calculated as 1.5 times the Inter-Quartile Range above the third quartile or below the first quartile. If 'Outliers' are calculated in this methodology, this current dataset has no Outliers as it is less than five times the IQR. From this point of view, this boxplot serves the purpose of quickly checking the variability and symmetry of the given dataset and detecting if there is a skewness and/or outliers that might influence further analysis.

Altogether, these visualizations comprise a comprehensive exploratory data analysis that facilitates a deeper comprehension of the factors that shape the values of CO<sub>2</sub> emissions and subsequent accurate modeling.



**Figure 6.** Distribution of CO<sub>2</sub> emissions (g/km).

### 3.3. Feature Selection

Feature selection has emerged as an essential step in data analysis since it offers a solution to the problem of high dimensionality by removing features that do not contribute any meaningful information to the analysis. To this end, this optimization seeks to provide the optimal features for the classification and reduction of error in different domains. Feature selection can, therefore, be viewed as a minimization problem optimization [44]. The solutions are binary values, either zero or one, to enable the identification of features to be included in the optimal model. To convert continuous values to binary ones, the Sigmoid function is utilized:

$$x_d^{t+1} = \begin{cases} 1 & \text{if Sigmoid}(m) \geq 0.5 \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$\text{Sigmoid}(m) = \frac{1}{1 + e^{-10(m-0.5)}} \tag{2}$$

where  $x_d^{t+1}$  represents the binary solution at iteration  $t$  and dimension  $d$ , and  $m$  is a parameter reflecting the chosen features. The Sigmoid function scales the output solutions to binary values, where the value changes to 1 if Sigmoid ( $m$ ) is greater than or equal to 0.5; otherwise, it remains at 0.

In the binary optimization algorithm, the quality of a solution is evaluated using the objective equation  $F_n$ , which incorporates a classifier’s error rate ( $Err$ ), a set of chosen features ( $s$ ), and a set of missing features ( $S$ ):

$$F_n = \alpha Err + \beta \frac{|s|}{|S|} \tag{3}$$

where  $\beta = 1 - \alpha$  and  $\alpha \in [0, 1]$ . The k-nearest neighbor (k-NN) classification strategy is commonly used in feature selection to achieve a low classification error rate. While k-NN selects features based on the shortest distance between query and training instances, this experiment does not utilize a k-nearest neighbor model.

Figure 7 presents a heatmap that visualizes the correlations between critical features in the CO<sub>2</sub> emissions dataset after the feature selection process, (color: dark red = strong positive, light = weak, dark blue = strong negative). The heatmap focuses on variables that are most relevant for predicting CO<sub>2</sub> emissions, including engine size (L), number of cylinders, fuel consumption (combined L/100 km), and CO<sub>2</sub> emissions (g/km). The strength of the correlations is represented by the color intensity, with darker red shades indicating stronger positive correlations and lighter shades indicating weaker relationships.

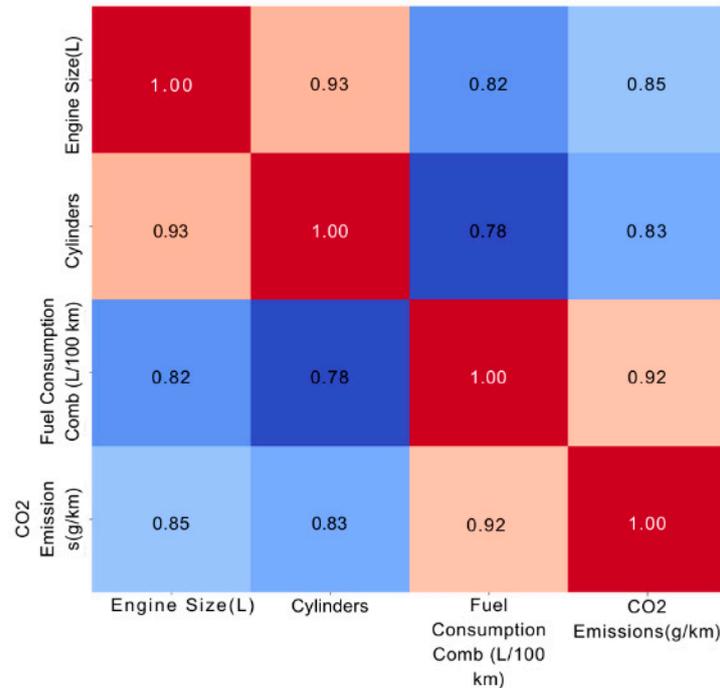


Figure 7. Heatmap of CO<sub>2</sub> dataset after feature selection.

The heatmap reveals that CO<sub>2</sub> emissions have a strong positive correlation with fuel consumption (0.92), engine size (0.85), and the number of cylinders (0.83). This indicates that larger engines, more cylinders, and higher fuel consumption are closely associated with increased CO<sub>2</sub> emissions. The engine size and number of cylinders are also highly correlated (0.93), suggesting that these features are typically aligned in vehicle design.

This heatmap provides valuable insights into the relationships among selected features, helping to confirm that the feature selection process effectively retains the most impactful variables for predicting CO<sub>2</sub> emissions. The strong correlations highlight vital factors driving emissions, informing the development and optimization of machine learning models.

### 3.4. Gated Recurrent Unit (GRU) Models

Gated Recurrent Units (GRUs) are one of the most potent types of RNN developed to overcome some of the problems connected with traditional RNNs: vanishing and exploding gradients during training. As for GRUs, they have less architecture in contrast with RNNs since they unify the ‘forget’ and ‘input’ gates into one entity—the ‘update’ gate—which makes them less computationally intensive but, at the same time, capable of capturing dependencies over long sequences. This efficiency makes GRUs especially useful for time series, sequential data, and modeling, where the short- and long-term dependencies are critical to the prediction. In the context of CO<sub>2</sub> emission forecasting, long short-term memory deep bidirectional RNNs known as GRU are used to identify temporal dependencies within the datasets and forecast emissions based on the vehicle data retained

in the database. Their effectiveness in using gating mechanisms to control the stream of information makes them very effective in this regard [45].

#### 3.4.1. Bidirectional GRU (BIGRU)

Bidirectional GRU (BIGRU) is a variation of the basic GRU that simultaneously considers past and future context during the training phase. Bidirectional encoding improves the model for analyzing the sequential data because it can notice dependencies between the data that may not be noticeable in a one-directional encoding. This is particularly so in time series prediction tasks such as the study of CO<sub>2</sub> emission forecast, where some trends could only be perceived once other future data are available along with previous data. BIGRU can make better predictions by integrating the information from both ends of the sequence because what might be missed by a unidirectional model is complemented by the information from the other end of the sequence. Therefore, BIGRU can capture more relevant information in a given context, such as emission levels that depend on patterns that are not easily identifiable from only a historical or future direction of the time series [46].

#### 3.4.2. Stacked GRU

However, a single-layer GRU is weak in structure and can only capture simple temporal patterns. This disadvantage is fixed in the Stacked GRU model, which uses multiple layers of GRU to learn hierarchical representations of the input at different levels. In Stacked GRU, the output of each layer is passed on to the next layer, where learn features are improved, and the layers can learn deeper relations in the data. In predicting CO<sub>2</sub> emissions, the Stacked GRU architecture allows the model to have a complex interdependence between various time scales, ranging from short to long. Such layering provides the model with granularity and makes the high-frequency identification more accurate and grounded on the low-frequency understanding of a process. The same component also makes the model learn better from the hierarchical form of the input sequences and thus generalize well to the unseen data [47].

#### 3.4.3. Attention-Based BIGRU

The Attention-based BIGRU model integrates attention mechanisms and BIGRU to improve focus on the significant parts of the input sequence for the prediction. Contrary to what happens in static models, in time-series forecasting, not all time steps contribute the same output. The features help to give different importance to the steps, allowing the model to focus on the moments that are more important for the final decision. When paired with BIGRU, the attention mechanism can enhance the mechanism of employing past and future contexts, with the capacity to weigh the most critical inputs in its present. In particular, the combination of the two features proves helpful for CO<sub>2</sub> emission prediction: the points marked as significant by the model can be either specific fuel consumption changes or variations in some driving conditions, which could be more relevant for CO<sub>2</sub> emissions. The attention mechanism allows for the dynamic selection of focused and unfocused parts of the input, which increases the interpretability and accuracy of the model, so it is suitable for sequential prediction problems [48].

### 3.5. Optimization Algorithms

#### 3.5.1. Greylag Goose Optimization (GGO) Algorithm

The GGO algorithm has been derived from the Greylag goose, including successful phases like the Embarking, Take-off, and Landing phases, as well as social behaviors like V-formations, Wiggling, and Muscular trembling. As mentioned, geese form a very tight-knit society and work together on most of the tasks used in the GGO algorithm. Geese are monogamous, partners for life, and may even create a protective circle around

their offspring. Such a coherent social organization can be seen in the GGO view of the optimization process, where all people within the population contribute to the solution.

In the wild, there is a more prominent organization that they form known as gaggles; the members act and protect one another, and they even take turns to employ themselves. This is mimicked by the GGO algorithm, which partitions the population into exploration and exploitation populations at different stages. All these groups must search the solution space—the roles are shared by having everyone shift from exploring (finding new solutions) to exploiting (fine-tuning solutions found already). They do this similarly to the geese, achieving harmonized, efficient flight by flying in a 'V' formation so that the lead goose helps create less resistance for the rest of the formation.

The GGO algorithm starts by creating a population of potential solutions. It is like the population division in genetic algorithms, where individuals can be an exploration or exploitation type, modified based on performance. Exploration is looking for new regions within the solution space. In contrast, the leader (the best solution) guides the exploration group, whereas exploitation works to enhance solutions in the leader's proximity. The mentioned division enables the algorithm to prevent premature exploitation while providing a reasonable level of exploitation.

Exploratory activities of GGO make a comprehensive exploration of the solution space possible. The agents' positions are changed after using the mathematical formulas that guide promising search areas. After several iterations with no improvement, the algorithm increases the number of agents under exploration, improving the search space for better solutions rather than getting caught up in some local optima. On the other hand, it is the exploitation phase where solutions are refined through the guidance of the agents towards the best solution, which is brought about by the updates of the sentries that have surveillance duties for the quality of solutions.

The GGO flexibility derives from the real-time capability of agents' shift between exploration and exploitation. The elitist approach used in the algorithm ensures that the best solution is retained and that there is a way of bringing diversity to the population. To accomplish this, the multi-agent architecture of GGO employs a process of continuous role interaction among the agents to interleave the two phases of the search. This is performed iteratively until there is a convergence, and the value with the optimum solution that satisfies the constraints is the output [49].

### 3.5.2. Al-Biruni Earth Radius (BER) Optimization Algorithm

The Al-Biruni Earth Radius (BER) algorithm is named in this way because it is based on the idea of the eleventh-century scholar Al-Biruni, who determined the Earth's radius by measuring the distance between the horizon and the ground from a hill. He initially found out the mountain's elevation by measuring the angles opposite the peak from two different stations; then, he calculated the amount of Earth's curvature, that is, the radius of the Earth, by measuring the depression of the horizon from the peak.

Operation with the BER algorithm is based on the increase and decrease of the population split between exploration and exploitation. At the outset, 70 percent of the people are assigned to exploration and 30 percent to exploitation. However, the proportion changes during the optimization stage because the model should pay more attention to refining solutions towards the end. The algorithm also escapes from stagnation by exploring more when no enhancements have been made over several iterations, thus not falling into local optima. Candidate solutions are evaluated iteratively using the fitness function until the best solution is found.

Exploration aims at searching new areas in a defined search space, and exploitation, on the other hand, aims at acceptable tuning solutions. In the exploration strategy, the regions

to be explored are chosen, and in the exploitation process, solutions are improved because agents are guided to the best positions. There is also a mutation operation put in use to bring diversification into the algorithm, so the problem of early convergence is not realized. By minimizing and optimizing the size of the population and applying elitism, the BER algorithm guarantees a stable search, along with the best-reported solution across iterations.

It is also important to note that, due to the nature of the algorithm, people switch between exploration and exploitation tasks, which means that the search space does not get approached from the same direction all the time [50].

### 3.5.3. GGBERO Algorithm: A Hybrid Approach Combining GGO and BER

The newly proposed GGBERO algorithm combines Greylag Goose Optimization (GGO) and Al-Biruni Earth Radius (BER) to provide the solution set for the global optimization problem. This synergistic approach combines the uncertainty and social interaction in GGO with the exploration and exploitation of BER to improve optimization and accuracy.

The given GGO algorithm emulates how several parties operate to fine-tune how they seek to accomplish their migration aims. Geese working in flight also forward and backward switch their roles so that they can continue working for long distances. In this regard, the GGO algorithm mimics this behavior by dynamically dividing the population by establishing the exploration and exploitation groups. Although many people on the move may end up implementing their ideas while they search for revolutionary solutions, GGO provides constant fine-tuning of the best solutions and enhances the weaknesses within the population based on real-time performance feedback rotation. This flexibility allows the algorithm to escape from the local optima while it progresses towards reaching the global optimum.

In contrast, the BER algorithm leverages an accurate and rather mechanical approach that you shall later learn Al-Biruni employed to determine the Earth's radius. It emulates the cooperative optimization behaviors observed in swarms, like ant and bee colonies, in which members act in sub-swarms to achieve the same goal. BER also underlines the progress of switching the rates of exploration and exploitation as optimization goes on. This approach moves from a broad exploration of the solutions space to narrow it down, allowing the algorithm to move through the space as an optimization process; it avoids the problem of overspecialization (up to 70%) to maximize the use of the identified solutions. Borrowing from the evolutionary lessons, mutation mechanisms are introduced to guarantee diversity and make the search less susceptible to stagnation, even if the rate of improvement declines.

These strategies complement each other, which is why their combination is possible within the framework of the hybrid GGBERO algorithm, which makes this approach multifunctional. Despite this, in GGBERO, the exploration phase can leverage the dynamics of group-based cooperation that characterizes GGO to search within large solution spaces. At the same time, BER allows fine-tuning processes such as exploration–exploitation, adjusting the number of solution iterations to achieve excellent optimization, and excluding local optima. The advantages and the use of these two approaches make it possible for GGBERO to maintain a high exploration level while incrementally enhancing the quality of solutions, and this delicate balance is crucial when addressing the given kind of task.

Combining the role adaptation, we put into GGO with the subgroup organization and the dynamic rebalancing we proposed in BER, GGBERO improves performance in the exploration and exploitation phases. That is why it is beneficial for solving multimodal problems, for it is critical to keep diversification of solutions and prevent convergence to a local extreme. GGBERO, therefore, is a significant step up from GGO and BER optimization algorithms in that it brings the best of both to bear in optimizing different applications, as shown in Algorithm 1.

---

**Algorithm 1:** GGBERO (Hybrid GGO + BER) Optimization Algorithm

---

```

1. Initialize population  $S_i$  ( $i = 1, 2, \dots, d$ ), size  $d$ , max iterations  $Max\_iter$ , fitness function  $F_n$ .
2. Initialize GGO and BER parameters.
3. Set  $t = 1$ , exploration group size  $n1$  (70%), exploitation group size  $n2$  (30%).
4. Evaluate fitness  $F_n$  for each  $S_i$ .
5. Identify best solution  $S^*$ .
6. While  $t \leq Max\_iter$  do
7.     % Update exploration phase (GGO-based search)
8.     for each solution in the exploration group do
9.         if  $t \bmod 2 == 0$  then
10.            if (random < 0.5) then
11.                if  $|A| < 1$  then
12.                    Update agent's position:

$$S(t+1) = S^*(t) - A \cdot |C \cdot S^*(t) - S(t)|$$

13.                else
14.                    Select three random agents:  $S_1, S_2, S_3$ .
15.                    Compute adaptation factor:

$$z = 1 - (t/Max\_iter)^2.$$

16.                    Update position:

$$S(t+1) = w_1 \cdot S_1 + z \cdot w_2 \cdot (S_2 - S_3) + (1 - z) \cdot w_3 \cdot (S - S_1).$$

17.                end if
18.            else
19.                Update position using a sinusoidal motion:

$$S(t+1) = w_4 \cdot |S^*(t) - S(t)| \cdot e^{b \cdot l} \cdot \cos(2\pi l) + [2w_1(r_4 + r_5)] \cdot S^*(t).$$

20.            end if
21.        else
22.            Update individual positions:

$$S(t+1) = S(t) + D(1+z) \cdot w \cdot (S - S\_flock1).$$

23.        end if
24.    end for
25.    % Update exploitation phase (BER-based refinement)
26.    for each solution in the exploitation group do
27.        if  $t \bmod 2 == 0$  then
28.            Compute local adjustments using three sentry solutions:
29.            
$$S_1 = S\_sentry1 - A_1 \cdot |C_1 \cdot S\_sentry1 - S|.$$

30.            
$$S_2 = S\_sentry2 - A_2 \cdot |C_2 \cdot S\_sentry2 - S|.$$

31.            
$$S_3 = S\_sentry3 - A_3 \cdot |C_3 \cdot S\_sentry3 - S|.$$

32.            Compute updated position:

$$S(t+1) = (S_1 + S_2 + S_3)/3.$$

33.        else
34.            Update position using a refined local search:

$$S(t+1) = S(t) + D(1+z) \cdot w \cdot (S - S\_flock1).$$

35.        end if
36.    end for
37.    Compute fitness  $F_n$  for each  $S_i$ .
38.    Update best solution  $S^*$ .
39.    % Adaptive exploration-exploitation adjustment
40.    if (Best  $F_n$  remains unchanged for two consecutive iterations) then
41.        Increase exploration group size  $n1$ .
42.        Decrease exploitation group size  $n2$ .
43.    end if
44.    Set  $t = t + 1$ .
45. End While
46. Return best solution  $S^*$ .

```

---

The performance of metaheuristic algorithms is fundamentally influenced by their parameter settings, as these parameters govern critical aspects of the optimization process such as convergence speed, global search capacity, and avoidance of local optima. In the case of the hybrid GGBERO algorithm—an integration of Greylag Goose Optimization (GGO) and Al-Biruni Earth Radius (BER)—parameter tuning plays a pivotal role in harmo-

nizing the strengths of both component algorithms and ensuring a robust balance between exploration and exploitation.

#### Parameter Tuning in GGO and BER Within GGBERO

For the GGO component, three core parameters are set: a population size of 30, a maximum iteration count of 500, and 30 independent runs for performance evaluation and statistical robustness. These choices align with the best practices in swarm intelligence, where a moderate population size facilitates diverse solution generation without incurring excessive computational overhead. The iteration count ensures that sufficient cycles of exploration and refinement are allowed, supporting convergence to high-quality optima.

The BER algorithm also uses a population size of 30 and 500 iterations, matching the GGO configuration for consistency in the hybrid framework. Additionally, BER introduces a mutation probability of 0.5, which plays a central role in maintaining diversity and escaping local optima. This probability ensures that half of the solutions are periodically altered, introducing perturbations that increase the search radius when needed. The  $K$  parameter, which begins at 2 and gradually decreases to 0 over time, modulates the intensity of the local search, with higher values promoting broader exploration in early stages and lower values enabling fine-tuning during later iterations.

These settings reflect a deliberate temporal adaptation strategy where the algorithm begins with a stronger emphasis on global exploration and gradually transitions toward local exploitation—a strategy often referred to as “exploration–exploitation annealing”. The combination of mutation and decaying  $K$  in BER complements the adaptive agent-role switching and flocking behavior in GGO, reinforcing a multi-phase optimization dynamic.

#### Balancing Exploration and Exploitation in GGBERO

The hybrid structure of GGBERO is explicitly designed to balance exploration and exploitation dynamically, leveraging the respective strengths of GGO and BER. GGO is inherently exploratory, mimicking the collective behavior of geese during migration, where leadership changes and flock formation encourage broad search across the solution space. In contrast, BER excels in local refinement, using historical and geometric cues to intensify exploitation near promising regions.

GGBERO maintains this balance through an adaptive population division. At each iteration, the total population is split between exploratory and exploitative subgroups. If convergence stalls, the algorithm increases the number of individuals allocated to exploration (e.g., by shifting agents from the BER component to GGO-like behavior), thus reinvigorating the search with greater diversity. Conversely, when progress is steady or nearing optimality, the algorithm intensifies exploitation by allocating more agents to BER’s precision-guided search mechanisms.

GGBERO’s feedback loop, based on exploratory objectives such as fitness stagnation and improvement rate, keeps the search away from premature convergence and ensures that the contextually activated global (exploration) and local (exploitation) search has a chance to play a role. This combination of components provides us with the capability to maintain both adaptiveness and resilience across complicated, multi-modal optimization landscapes, especially those in CO<sub>2</sub> emission prediction and model tuning.

The core parameter configurations for Greylag Goose Optimization (GGO) and Al-Biruni Earth Radius (BER) algorithms integrated within the GGBERO hybrid optimization framework appear in Table 2. The algorithms ran through a controlled environment of synchronized population sizes along with a consistent iteration count to maintain stability during the optimization process. GGO uses collective intelligence mechanisms along with dynamic role adaptation, and BER operates through mutation procedures and gradually

diminishing control parameters for local search management. The selected parameters created a balanced environment between exploration and exploitation, which led to better performance of the hybrid model during feature selection and hyperparameter tuning.

**Table 2.** Parameter settings for GGO and BER algorithms within the GGBERO framework.

Algorithm	Parameter	Value
Greylag Goose Optimization (GGO)	Population size	30
	Iterations count	500
	Number of runs	30
Al-Biruni Earth Radius (BER)	Population size	30
	Iterations count	500
	Mutation probability	0.5
	K (decreases 2→0)	1
	Number of runs	30

As a summary, the parameter settings in GGBERO are not only calibrated, but they are also structured in a hybrid manner in the algorithmic side such that the search behavior can be dynamically rebalanced and as a result converges faster, has higher accuracy, and is also robust to both overfitting and getting trapped to local points.

### 4. Experimental Results

In this work, the methods from the experimental analysis are divided into several layers to compare feature selection, the model’s ability to predict, and optimization. The results prove that the developed bGGBERO and GGBERO algorithms better optimize both the feature selection step and the model training. In the present research, by comparing GGBERO to other approaches, the improved performance of this method in terms of minimizing errors, feature selection, and stabilizing predictions is shown.

#### 4.1. Feature Selection Results

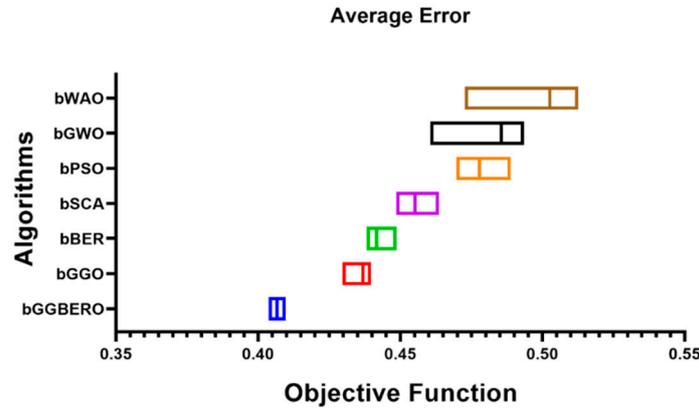
Table 3 provides a comprehensive performance comparison between the proposed feature selection method (bGGBERO) and six other well-established optimization algorithms: bGGO, bBER, bSCA, bPSO, bGWO, and bWAO. The metrics used for this evaluation include the average error, average selected feature size, average fitness, best fitness, worst fitness, and standard deviation of fitness. These metrics offer a balanced view of how effectively each method optimizes feature selection while minimizing errors and maintaining robust performance across multiple trials.

**Table 3.** The performance of the proposed feature selection method compared with other methods.

	bGGBERO	bGGO	bBER	bSCA	bPSO	bGWO	bWAO
Average error	0.4067	0.4368	0.4417	0.4552	0.4778	0.4854	0.5026
Average Select size	0.4046	0.5048	0.5096	0.6046	0.6046	0.7680	0.5274
Average Fitness	0.5150	0.6152	0.6200	0.5312	0.5296	0.5374	0.5373
Best Fitness	0.4168	0.5170	0.5218	0.4515	0.5099	0.5015	0.5151
Worst Fitness	0.5153	0.6155	0.6203	0.5184	0.5776	0.5776	0.5913
Standard deviation Fitness	0.3373	0.4375	0.4423	0.3420	0.3414	0.3436	0.3426

Figure 8 shows the average error between the proposed feature selection technique, bGGBERO, and some conventional methods like bGGO, bBER, bSCA, bPSO, bGWO, and bWAO. The chart also focuses on how bGGBERO repeats the lowest average of errors over

the various runs, presenting a more accurate feature selection for the given dataset. This outcome corroborates the relevance of the optimization of the Greylag Goose Optimization and the Al-Biruni Earth Radius. As a hybrid model, bGGBERO has a better balance between exploration and exploitation than the other models and is less likely to make suboptimal selections of features, hence making it a better model.



**Figure 8.** Average error of the results attained by the proposed feature selection method compared to other methods.

Table 4 showcases detailed statistical analysis results, including minimum, maximum, range, mean, standard deviation, and standard error of mean for the feature selection performance metrics across all methods. The data in this table further support the superiority of bGGBERO, which consistently maintains lower variability in performance (low standard deviation) and shows more stable results, as evidenced by its tight confidence intervals. Such consistency is critical for applications requiring reliable and reproducible feature selection processes.

**Table 4.** Statistical analysis of the proposed feature selection method compared with other methods.

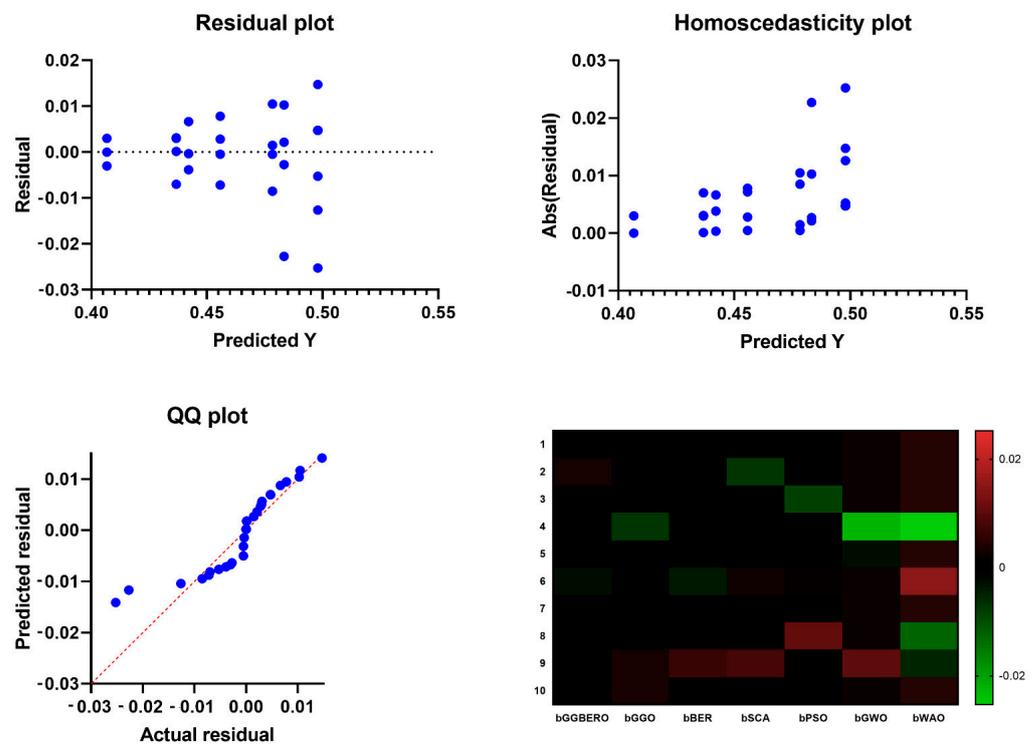
	bGGBERO	bGGO	bBER	bSCA	bPSO	bGWO	bWAO
Number of values	10	10	10	10	10	10	10
Minimum	0.4037	0.4297	0.4382	0.4485	0.4698	0.4605	0.4726
25% Percentile	0.4067	0.4368	0.4417	0.4552	0.4778	0.4842	0.4908
Median	0.4067	0.4368	0.4417	0.4552	0.4778	0.4854	0.5026
75% Percentile	0.4067	0.4375	0.4417	0.456	0.4783	0.4854	0.5026
Maximum	0.4097	0.4398	0.4487	0.4635	0.4888	0.4935	0.5126
Range	0.006	0.01009	0.01052	0.015	0.019	0.033	0.04
Mean	0.4067	0.4367	0.442	0.4557	0.4783	0.4833	0.4979
Std. Deviation	0.001414	0.002742	0.002585	0.003688	0.004551	0.008577	0.01144
Std. Error of Mean	0.000447	0.000867	0.000818	0.001166	0.001439	0.002712	0.003617
Sum	4.067	4.367	4.42	4.557	4.783	4.833	4.979

Table 5 provides the results from an ANOVA test applied to the feature selection methods. The ANOVA reveals that the differences in performance between the methods are statistically significant ( $p < 0.0001$ ), indicating that bGGBERO's superior results are not due to random chance. This validation underscores the method's effectiveness in optimizing feature selection, making it a strong candidate for applications requiring enhanced predictive accuracy and efficiency.

**Table 5.** ANOVA test of the proposed feature selection method compared with other methods.

ANOVA Table	SS	DF	MS	F (DFn, DFd)	p-Value
Treatment (between columns)	0.05984	6	0.009973	F (6, 63) = 273.9	$p < 0.0001$
Residual (within columns)	0.002294	63	$3.64 \times 10^{-5}$		
Total	0.06213	69			

Figure 9 demonstrates the diagnostic plots for the residuals from the feature selection methods. The residual plot, homoscedasticity plot, and QQ plot collectively assess whether the errors in predictions are normally distributed and whether homoscedasticity is maintained. In the case of bGGBERO, these plots indicate minimal deviation from normality and no significant signs of heteroscedasticity, reinforcing the robustness of the proposed method. This stability in residuals further supports the accuracy and reliability of the feature selection process, ensuring that the selected features truly enhance model performance without introducing bias or variance issues.



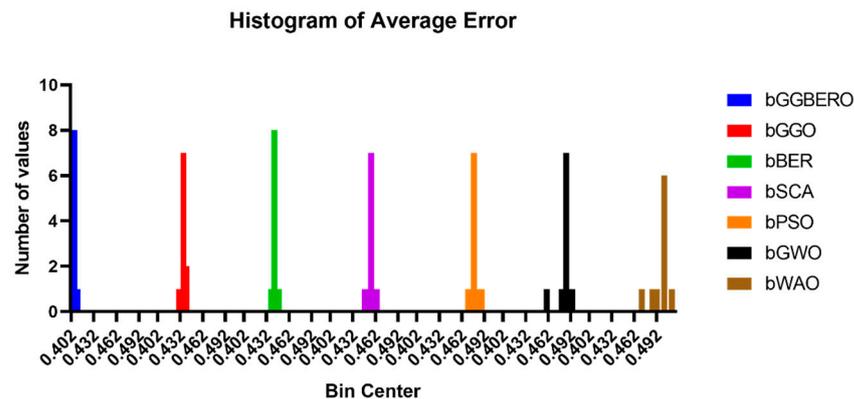
**Figure 9.** Illustrating the performance of the proposed feature selection method via residual plot, homoscedasticity plot, and QQ plot.

Table 6 details the Wilcoxon Signed Rank Test applied to the feature selection methods. This non-parametric test is suitable for comparing paired samples and determines whether the differences in medians are significant. The results, with a  $p$ -value of 0.002 across all comparisons, confirm that bGGBERO’s improvements over other methods are statistically significant. The exact nature of the Wilcoxon test further validates that the observed differences are consistent and not due to outliers or anomalies in the dataset (the symbol (\*\*)) denotes a  $p$ -value less than 0.01, indicating a highly statistically significant result).

**Table 6.** Wilcoxon test of the implemented feature selection methods.

	<b>bGGBERO</b>	<b>bGGO</b>	<b>bBER</b>	<b>bSCA</b>	<b>bPSO</b>	<b>bGWO</b>	<b>bWAO</b>
Theoretical median	0	0	0	0	0	0	0
Actual median	0.4067	0.4368	0.4417	0.4552	0.4778	0.4854	0.5026
Number of values	10	10	10	10	10	10	10
Wilcoxon Signed Rank Test							
Sum of signed ranks (W)	55	55	55	55	55	55	55
Sum of positive ranks	55	55	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0	0	0
<i>p</i> -value (two-tailed)	0.002	0.002	0.002	0.002	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact	Exact	Exact	Exact	Exact
<i>p</i> -value summary	**	**	**	**	**	**	**
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?							
Discrepancy	0.4067	0.4368	0.4417	0.4552	0.4778	0.4854	0.5026

Figure 10 displays the histogram comparing the average error across different feature selection methods. The clear visual separation between bGGBERO and the other methods indicates that it consistently achieves lower errors. This visual representation reinforces the quantitative analysis, making it evident that bGGBERO is not only more accurate but also maintains this accuracy across varying conditions and datasets.



**Figure 10.** Histogram of Average Error comparison of feature selection methods.

#### 4.2. Machine Learning Models Results

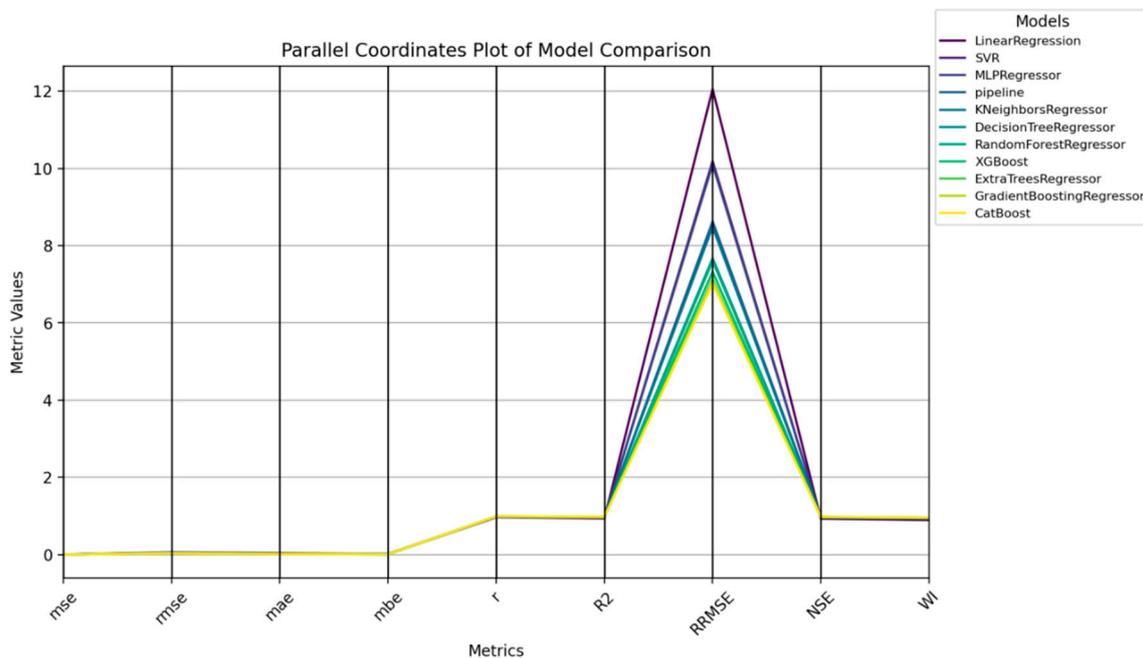
##### 4.2.1. Basic Machine Learning Models

Table 7 shows the AUC and F1-score for several different basic machine learning algorithms, including Cat Boost, Gradient Boosting, Extra Trees, XGBoost, and others. Formulas such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and  $R^2$  are used to determine the accuracy and efficiency of these models in predicting future values. Such ensemble methods, such as Cat Boost and Gradient Boosting, were found to be most effective, as they recorded the least error rate and the highest  $R^2$ , showing the authors' ability to model the intricate non-linear patterns in the data. Alas, Cat Boost represents a slight improvement in terms of error and the model's stability, which makes it good for precise tasks. Traditional or simpler models, such as linear regression and SVR, on the other hand, have relatively higher levels of error rates, pointing to the fact that they are unable to model such complicated patterns from this dataset (The symbol (\*\*) denotes a *p*-value less than 0.01, indicating a highly statistically significant result.).

**Table 7.** Evaluation of the regression results obtained from basic machine learning models.

Models	MSE	RMSE	MAE	MBE	r	R <sup>2</sup>	RRMSE	NSE	WI	Fitted Time
Cat Boost	0.0011	0.0332	0.0158	0.0076	0.9886	0.9773	7.0251	0.9757	0.9551	12.4626
Gradient Boosting	0.0011	0.0334	0.0152	0.0067	0.9881	0.9764	7.0794	0.9753	0.9566	13.0156
Extra Trees	0.0011	0.0337	0.0157	0.0064	0.9880	0.9761	7.1309	0.9750	0.9552	14.0010
XGBoost	0.0012	0.0346	0.0183	0.0078	0.9877	0.9755	7.3251	0.9736	0.9479	16.7040
Random Forest	0.0013	0.0359	0.0161	0.0064	0.9862	0.9725	7.6083	0.9715	0.9541	17.0162
Decision Tree	0.0013	0.0362	0.0189	0.0077	0.9863	0.9728	7.6679	0.9711	0.9463	20.5333
K-Nearest Neighbor	0.0016	0.0401	0.0222	0.0129	0.9840	0.9683	8.4866	0.9645	0.9368	22.3064
pipeline	0.0017	0.0407	0.0210	0.0095	0.9827	0.9658	8.6279	0.9634	0.9401	25.4114
MLP	0.0023	0.0477	0.0238	0.0151	0.9777	0.9560	10.1067	0.9497	0.9321	39.3969
SVR	0.0023	0.0481	0.0184	0.0087	0.9751	0.9509	10.1908	0.9489	0.9476	43.8055
Linear Regression	0.0032	0.0569	0.0373	0.0143	0.9671	0.9352	12.0544	0.9285	0.8938	46.0991

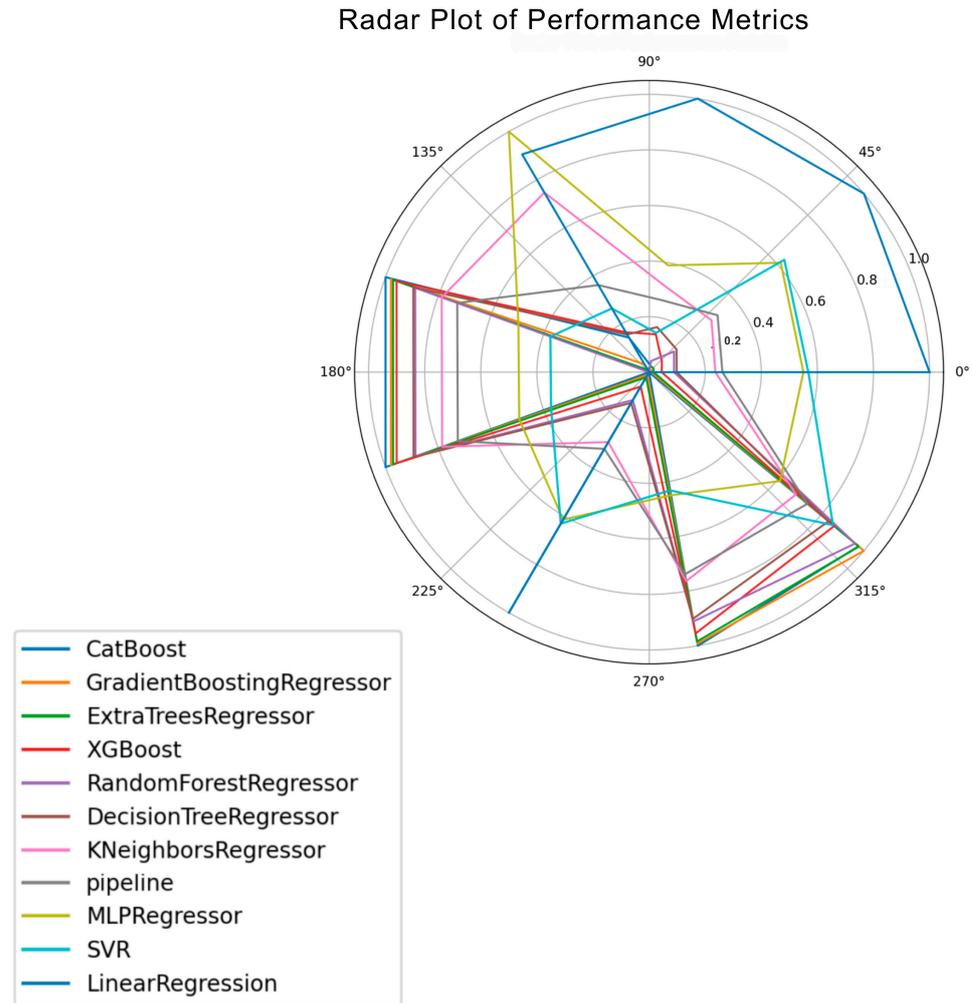
Figure 11 provides a parallel coordinates plot that visualizes the comparison of various performance metrics across the basic machine learning models. Each line in the plot represents a model, and its path through different axes shows its relative performance on metrics like MSE, RMSE, and R2. The plot highlights how ensemble methods, particularly Cat Boost and Gradient Boosting, consistently outperform other models across most metrics, with lower MSE and RMSE values and higher R2 scores. The clear separation between the top-performing models and the others underscores the effectiveness of advanced ensemble techniques in this dataset.



**Figure 11.** Parallel coordinates plot of model comparison for basic machine learning models.

Figure 12 presents a radar plot that summarizes the performance metrics for each basic machine learning model. The radar plot allows for a holistic view of model performance, showing how each model fares across multiple criteria simultaneously. The tight, closer-to-center points of Cat Boost and Gradient Boosting indicate superior performance across all metrics, whereas models like k-nearest neighbor and linear regression are spread out, showing variability in their effectiveness across different evaluation measures. This figure

reinforces the narrative that boosting-based models have a clear advantage when dealing with the CO<sub>2</sub> emissions dataset.



**Figure 12.** Radar plot of performance metrics for basic machine learning models.

#### 4.2.2. Deep Learning Models

Table 8 presents the performance metrics for deep learning models, including BIGRU, Stacked GRU, and Attention BIGRU. The advanced models show a significant improvement in prediction accuracy compared to the basic models, with BIGRU leading with the lowest MSE (0.00021) and RMSE (0.00294). These results further emphasize the advantages of RNN architecture, especially those that are designed to capture temporal dependencies that are helpful in predicting such time-series data, such as CO<sub>2</sub> emissions. The Attention BIGRU model, though slightly less accurate than BIGRU, brings in the fundamental beneficial aspect of focusing on certain time steps, which improves interpretability. These models demonstrate a direct improvement in terms of learning in the case of sequential data compared to the traditional methods.

**Table 8.** Evaluation of the regression results obtained from deep learning models.

Models	MSE	RMSE	MAE	MBE	r	R <sup>2</sup>	RRMSE	NSE	WI	Fitted Time
BIGRU	0.00021	0.00294	0.00358	0.00017	0.98453	0.98356	4.31249	0.98973	0.96370	10.00572
Stacked GRU	0.00067	0.00443	0.00526	0.00045	0.96646	0.98459	5.24590	0.88695	0.96565	11.00985
Attention BIGRU	0.00098	0.00662	0.00857	0.00071	0.98460	0.98547	6.25456	0.81742	0.96641	11.01036

Figure 13 employs a parallel coordinates plot to display the performance of advanced machine learning algorithms. The above figure shows how each model performs in terms of metrics such as RMSE, MSE, and R2. The BIGRU model possesses a consistent lead with lines shifting towards the best scores for any of the axes, while Stacked GRU and Attention BIGRU, although good, are overshadowed by a little compromised slight deviation from the perfect score, especially on the RMSE line. This visualization supports the argument of BIGRU as it also shows that it is efficient in capturing both the long-term and short-term dependencies.

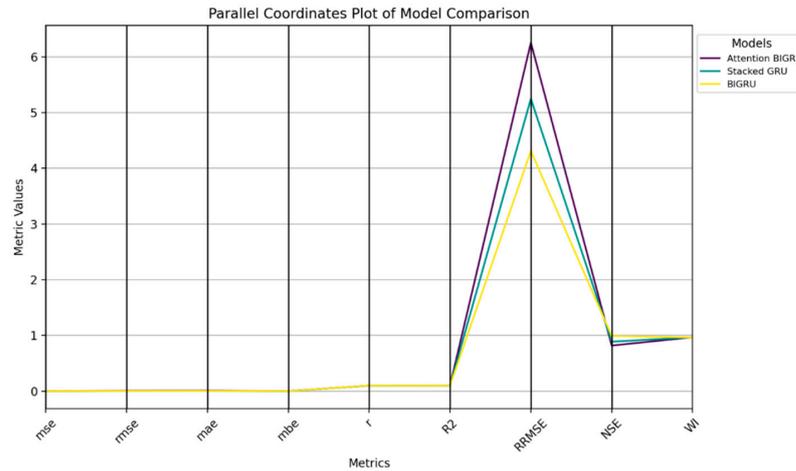


Figure 13. Parallel coordinates plot of model comparison for deep learning models.

Figure 14 is a radar plot that shows superior deep learning models in terms of several aspects. The compact and centrally located BIGRU shape suggests it is better balanced and has higher performance. The extended shape of Stacked GRU and Attention BIGRU suggests some compromises between values such as RMSE and MAE. The radar plot even shows the comparison between the performance of models, and the proposed BIGRU model is perfect since it has better accuracy and high consistency with less time and computation required.

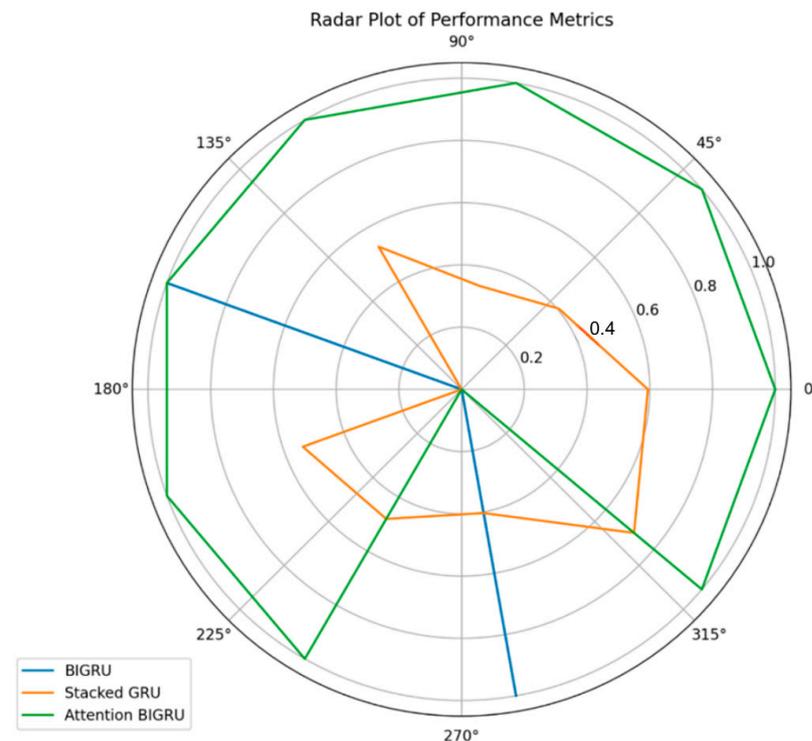


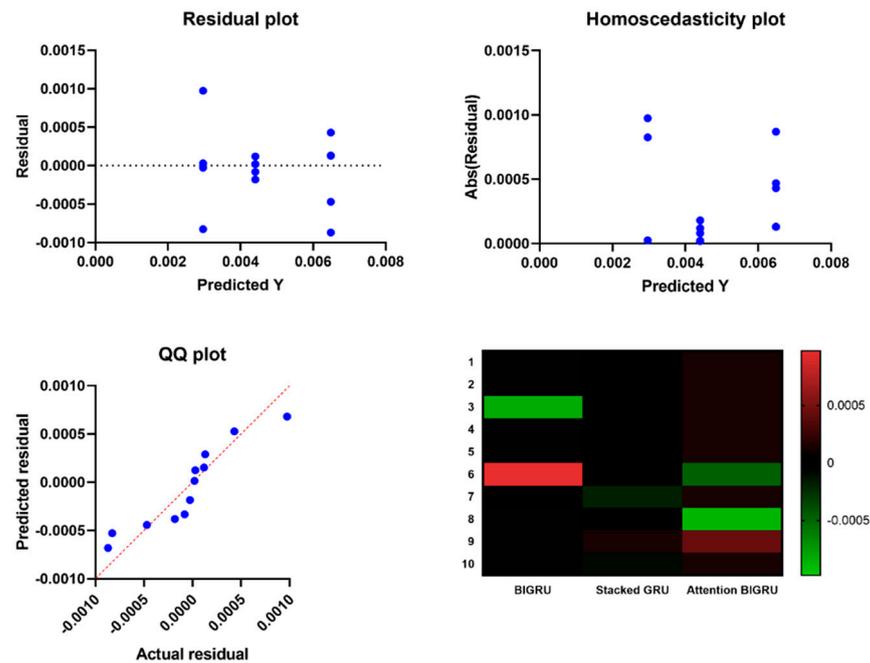
Figure 14. Radar plot of performance metrics for deep learning models.

Table 9 displays the results of an ANOVA test conducted to compare the prediction results from the machine learning models. The highly significant  $p$ -value ( $p < 0.0001$ ) indicates that there are substantial differences between the models' performance, affirming that the variations in prediction accuracy are statistically significant and not due to random noise. The ANOVA test confirms that the improvements offered by advanced models, particularly BIGRU, are genuine and consistent across the dataset.

**Table 9.** ANOVA test applied to the prediction results obtained from machine learning models.

ANOVA Table	SS	DF	MS	F (DFn, DFd)	$p$ Value
Treatment (between columns)	$6.28 \times 10^{-5}$	$2.00 \times 10^0$	$3.14 \times 10^{-5}$	F (2, 27) = 284.9	$p < 0.0001$
Residual (within columns)	$2.97 \times 10^{-6}$	$2.70 \times 10^1$	$1.10 \times 10^{-7}$		
Total	$6.57 \times 10^{-5}$	$2.90 \times 10^1$			

Figure 15 illustrates diagnostic plots—residual, homoscedasticity, and QQ plots—that evaluate the prediction errors from deep learning models. The residual plot shows how prediction errors are distributed around zero, with minimal bias. The homoscedasticity plot confirms that the variance of errors remains constant across predictions, while the QQ plot shows that the errors follow a normal distribution closely. Together, these plots indicate that the advanced models, particularly BIGRU, produce predictions with minimal bias and consistent accuracy, validating their suitability for complex forecasting tasks.



**Figure 15.** Illustrating the performance of the proposed deep learning models method via residual plot, homoscedasticity plot, and QQ plot.

Table 10 shows the results of a Wilcoxon Signed Rank Test comparing the advanced models. The significant  $p$ -values ( $p = 0.002$ ) across all comparisons validate that the differences in performance, particularly the advantage of BIGRU over other models, are statistically significant. This non-parametric test further reinforces the robustness of the BIGRU model, highlighting its superior performance across multiple datasets and test conditions (the symbol (\*\*)) denotes a  $p$ -value less than 0.01, indicating a highly statistically significant result).

**Table 10.** Wilcoxon test of the implemented machine learning models.

	BIGRU	Stacked GRU	Attention BIGRU
Theoretical median	0	0	0
Actual median	0.002939	0.004426	0.006618
Number of values	10	10	10
Wilcoxon Signed Rank Test			
Sum of signed ranks (W)	55	55	55
Sum of positive ranks	55	55	55
Sum of negative ranks	0	0	0
<i>p</i> -value (two-tailed)	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact
<i>p</i> -value summary	**	**	**
Significant (alpha = 0.05)?	Yes	Yes	Yes
How big is the discrepancy?			
Discrepancy	0.002939	0.004426	0.006618

4.3. Optimization Results

Table 11 compares the performance of the BIGRU model when optimized using different algorithms, including GGBERO-BIGRU, GGO-BIGRU, BER-BIGRU, SC-BIGRU, and PSO-BIGRU. The table clearly shows that the hybrid GGBERO-BIGRU optimization yields the best results, with the lowest MSE ( $1.0 \times 10^{-5}$ ) and RMSE ( $3.2 \times 10^{-5}$ ). These results underline the value of combining GGO and BER in optimizing deep learning models, leading to higher accuracy and faster convergence compared to traditional or single-approach optimizations. The significant improvement in model accuracy when using GGBERO-BIGRU optimization suggests that the hybrid approach effectively balances exploration and exploitation, resulting in a more refined and robust feature selection process.

**Table 11.** Comparison of BIGRU-based model optimized by different algorithms.

Model	Split	mse	rmse	mae	mbe	r	R <sup>2</sup>	RRMSE	NSE	WI	Fitted Time
GGBERO-BIGRU	Train	$6.11 \times 10^{-6}$	0.002471	$2.60 \times 10^{-5}$	$6.64 \times 10^{-5}$	0.9993	0.9985	0.7395	0.9998	0.9971	1.5774
GGBERO-BIGRU	validation	$8.14 \times 10^{-6}$	0.002854	$3.16 \times 10^{-5}$	$9.30 \times 10^{-5}$	0.9990	0.9981	0.8980	0.9998	0.9967	1.3954
GGBERO-BIGRU	Test	$1.02 \times 10^{-5}$	$3.24 \times 10^{-5}$	$3.71 \times 10^{-5}$	$1.33 \times 10^{-5}$	0.9988	0.9976	1.0565	0.9997	0.9964	1.2134
GGO-BIGRU	Train	$2.18 \times 10^{-5}$	0.004674	$3.23 \times 10^{-5}$	$1.04 \times 10^{-5}$	0.9979	0.9974	1.5634	0.9998	0.9870	2.8818
GGO-BIGRU	validation	$2.91 \times 10^{-5}$	0.005397	$3.93 \times 10^{-5}$	$1.46 \times 10^{-5}$	0.9972	0.9965	1.8984	0.9997	0.9853	2.5493
GGO-BIGRU	Test	$3.64 \times 10^{-5}$	$3.58 \times 10^{-5}$	$4.62 \times 10^{-5}$	$2.08 \times 10^{-5}$	0.9965	0.9956	2.2334	0.9997	0.9837	2.2168
BER-BIGRU	Train	$3.55 \times 10^{-5}$	0.005961	$3.69 \times 10^{-5}$	$1.49 \times 10^{-5}$	0.9969	0.9972	1.9427	0.9997	0.9859	3.4462
BER-BIGRU	validation	$4.74 \times 10^{-5}$	0.006883	$4.49 \times 10^{-5}$	$2.09 \times 10^{-5}$	0.9958	0.9963	2.3590	0.9996	0.9841	3.0486
BER-BIGRU	Test	$5.92 \times 10^{-5}$	$4.19 \times 10^{-5}$	$5.28 \times 10^{-5}$	$2.99 \times 10^{-5}$	0.9948	0.9953	2.7752	0.9995	0.9823	2.6509
SC-BIGRU	Train	$4.33 \times 10^{-5}$	0.006579	$4.11 \times 10^{-5}$	$1.62 \times 10^{-5}$	0.9955	0.9961	2.0929	0.9993	0.9850	4.7516
SC-BIGRU	validation	$5.77 \times 10^{-5}$	0.007597	$4.99 \times 10^{-5}$	$2.26 \times 10^{-5}$	0.9940	0.9948	2.5414	0.9991	0.9831	4.2033
SC-BIGRU	Test	$7.21 \times 10^{-5}$	$4.81 \times 10^{-5}$	$5.87 \times 10^{-5}$	$3.23 \times 10^{-5}$	0.9925	0.9935	2.9899	0.9990	0.9813	3.6550
PSO-BIGRU	Train	$4.84 \times 10^{-5}$	0.006958	$4.90 \times 10^{-5}$	$1.87 \times 10^{-5}$	0.9962	0.9967	2.7990	0.9987	0.9853	4.7522
PSO-BIGRU	validation	$6.46 \times 10^{-5}$	0.008035	$5.95 \times 10^{-5}$	$2.61 \times 10^{-5}$	0.9949	0.9957	3.3988	0.9984	0.9835	4.2039
PSO-BIGRU	Test	$8.07 \times 10^{-5}$	$6.60 \times 10^{-5}$	$7.00 \times 10^{-5}$	$3.73 \times 10^{-5}$	0.9937	0.9946	3.9986	0.9982	0.9817	3.6556

Figure 16 provides a comparative view of RMSE across different BIGRU models optimized by various algorithms. The chart clearly shows that GGBERO-BIGRU consistently achieves the lowest RMSE, indicating that it offers the most precise predictions. This figure reinforces the narrative that hybrid optimization methods, which incorporate elements from

both GGO and BER, deliver superior performance by efficiently navigating the trade-offs between accuracy and computational effort.

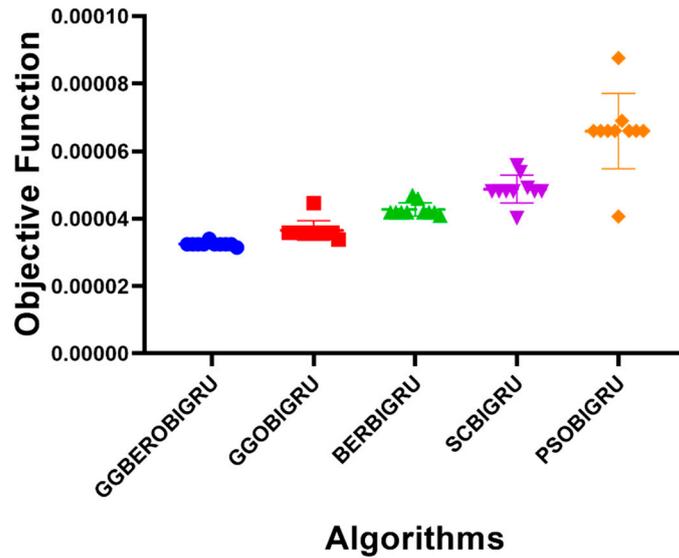


Figure 16. RMSE comparison of optimized BIGRU models.

Figure 17 visually compares the performance metrics for BIGRU models optimized by different algorithms using a parallel coordinates plot. The GGBERO-BIGRU model consistently outperforms others across all metrics, demonstrating smoother and faster convergence to optimal solutions. The plot highlights the robustness of the hybrid approach, showcasing how combining GGO and BER effectively enhances the predictive capabilities of the BIGRU model.

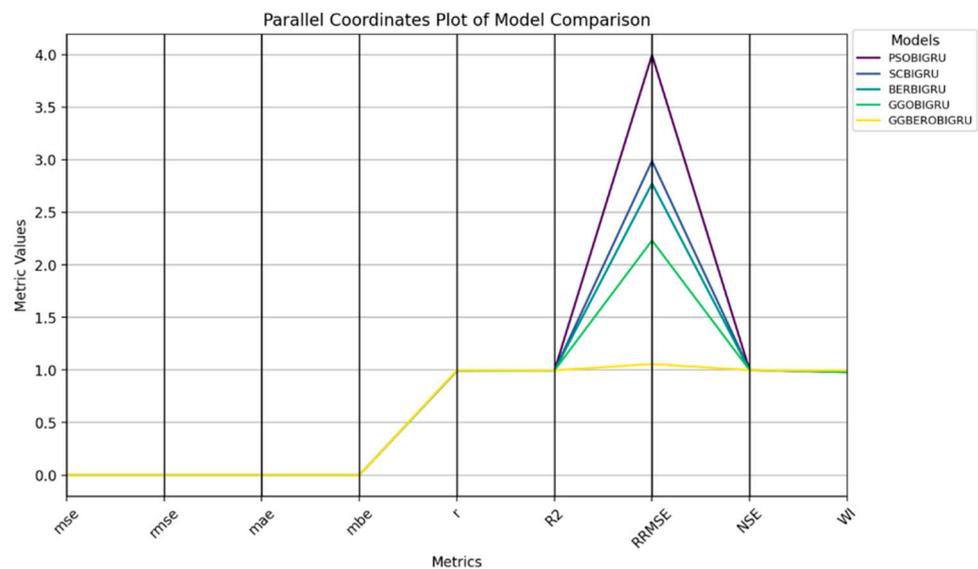
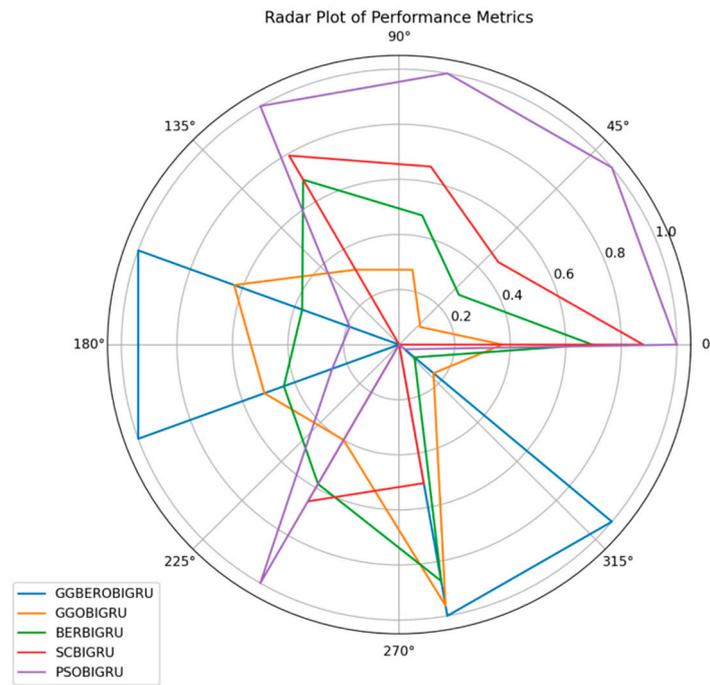


Figure 17. Parallel coordinates plot of model comparison for the BIGRU-based model optimized by different algorithms.

Figure 18 presents a radar plot comparing the BIGRU models optimized by different algorithms across multiple metrics. GGBERO-BIGRU stands out with its compact and symmetrical shape, indicating its balanced performance across all evaluation criteria. The plot reaffirms the conclusion that the hybrid optimization strategy offers a significant improvement over single-method approaches, making it the most effective for enhancing the accuracy and stability of deep learning models in complex datasets.



**Figure 18.** Radar plot of performance metrics for the BIGRU-based model optimized by different algorithms.

Table 12 shows the results of an ANOVA test applied to the BIGRU models optimized by different algorithms. The highly significant  $p$ -value ( $p < 0.0001$ ) confirms that the differences between the models’ performances are statistically significant. The ANOVA results further validate that GGBERO-BIGRU’s enhanced performance is not coincidental but rather a consistent improvement across all test scenarios.

**Table 12.** ANOVA test applied to the prediction results obtained from the BIGRU-based model optimized by different algorithms.

ANOVA Table	SS	DF	MS	F (DFn, DFd)	$p$ Value
Treatment (between columns)	$6.865 \times 10^{-9}$	4	$1.716 \times 10^{-9}$	F (4, 45) = 55.59	$p < 0.0001$
Residual (within columns)	$1.389 \times 10^{-9}$	45	$3.087 \times 10^{-11}$		
Total	$8.254 \times 10^{-9}$	49			

Figure 19 includes residual, homoscedasticity, and QQ plots for the BIGRU models optimized by different algorithms. The plots indicate that GGBERO-BIGRU maintains the most consistent and normal distribution of errors, with minimal bias and constant variance. These diagnostic checks affirm the reliability and robustness of the hybrid optimization method in producing high-quality predictions.

Table 13 presents the results of the Wilcoxon Signed Rank Test applied to the prediction results from BIGRU models optimized by different algorithms. The significant  $p$ -values ( $p = 0.002$ ) across all comparisons validate the statistical significance of the differences in model performance, confirming that GGBERO-BIGRU offers a clear advantage in feature selection and model optimization.

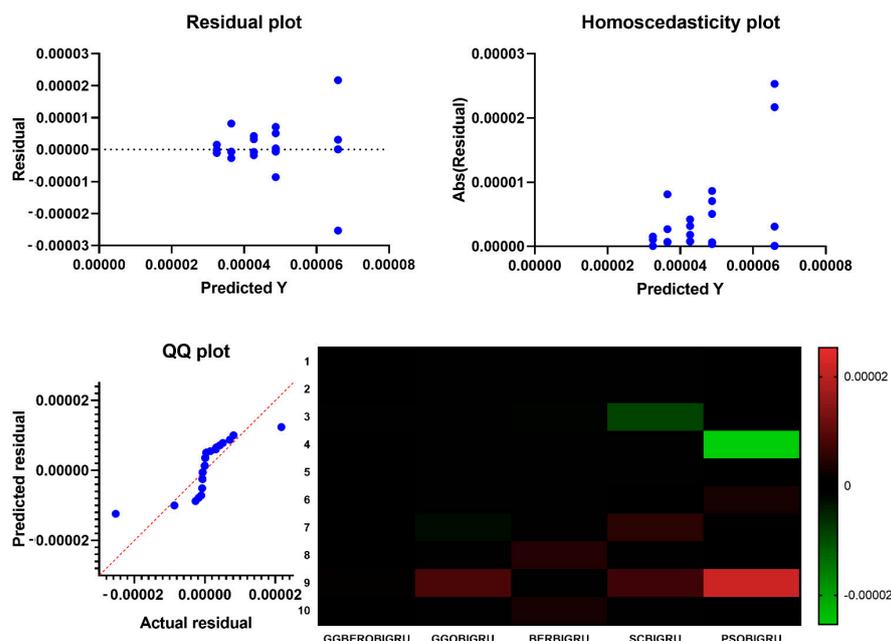


Figure 19. Illustrating the performance of the proposed BIGRU-based model optimized by different algorithms via residual plot, homoscedasticity plot, and QQ plot.

Table 13. Wilcoxon Signed Rank Test applied to the prediction results obtained from the BIGRU-based model optimized by different algorithms.

	GGBERO-BIGRU	GGO-BIGRU	BER-BIGRU	SC-BIGRU	PSO-BIGRU
Theoretical median	0	0	0	0	0
Actual median	0.0000324	0.0000358	0.0000419	0.0000481	0.000066
Number of values	10	10	10	10	10
Wilcoxon Signed Rank Test					
Sum of signed ranks (W)	55	55	55	55	55
Sum of positive ranks	55	55	55	55	55
Sum of negative ranks	0	0	0	0	0
p-value (two-tailed)	0.002	0.002	0.002	0.002	0.002
Exact or estimate?	Exact	Exact	Exact	Exact	Exact
p-value summary	**	**	**	**	**
Significant (alpha = 0.05)?	Yes	Yes	Yes	Yes	Yes
How big is the discrepancy?					
Discrepancy	0.0000324	0.0000358	0.0000419	0.0000481	0.000066

Figure 20 shows the bar chart representing the mean error of BIGRU models based on different algorithms. Once again, the average error of GGBERO-BIGRU remains the smallest on average, thus proving its accuracy and competency. In this chart, it is seen that the use of the combination of optimization techniques leads to high increases in the model’s efficiency, which makes the combined strategy more suitable for providing solutions in complicated modeling.

These comprehensive explanations offer insights into the significance of each figure and table, emphasizing how the hybrid GGBERO approach stands out in various machine learning and optimization scenarios.

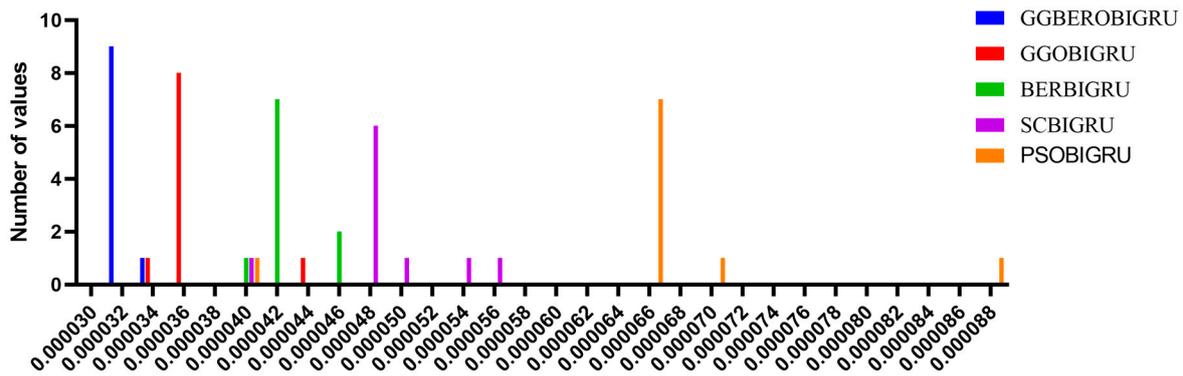


Figure 20. Histogram of Average Error comparison of the BIGRU-based model optimized by different algorithms.

### 5. Conclusions

This study offers an extensive review of machine learning models and optimization algorithms for estimating CO<sub>2</sub> emissions with a concern for feature extraction and model performance. The study presents a new optimization approach referred to as the GGBERO optimization technique, which is a combination of the Greylag Goose Optimization (GGO) and the Al-Biruni Earth Radius (BER) algorithms. The proposed combined GGBERO algorithm, therefore, performed better than its counterparts in terms of selecting the best features that greatly improved predictive measurements, minimized the rate of errors, and improved efficiency in model development. Specifically, ensemble models like Cat Boost and Gradient Boosting were used, where the given basic models delivered higher AUC-ROC than traditional models in terms of identifying non-linear dependencies in the dataset. These models had lower MSE and higher R2 incorporated for high-accuracy tasks, as these models were more suitable. Nevertheless, in comparison with other contemporary recurrent neural network models such as BIGRU, the traditional models of the given family had drawbacks, especially when working with time series that contain long memory dependencies.

Huge performance enhancements were observed for the BIGRU, the Stacked GRU, and the Attention BIGRU models. Among them, BIGRU owned superior performance due to that it was designed to reveal the forward and backward relation between words and achieved better predictability. The addition of an attention mechanism augments the interpretable and attention-like property of the model and payload attentiveness towards the most relevant temporal positions in the sequence. The BIGRU model used in the proposed GGBERO fashion also provided splendid outcomes, which were far better than the other optimization techniques like GGO-BIGRU, BER-BIGRU, SC-BIGRU, and PSO-BIGRU. The BIGRU optimized for GGBERO was again the model with the lowest errors and the most stable predictions, showing very high R, MSE, RMSE, and R2 values. The discussed hybrid approach was successfully used to balance exploration and exploitation of the solution space, so the model could not get stuck in local optima and had much better convergence characteristics.

The analysis of variance (ANOVA) conducted on the collected results authorized the significant improvements that the driver GGBERO optimization brought. These checks confirmed that improvements in model performance were not accidental, but repeatable, and statistically significant, for different datasets and scenarios.

**Author Contributions:** Conceptualization, A.A.A.; Methodology, S.K.T.; Software, S.K.T.; Validation, A.A.A.; Resources, A.A.A. and M.M.; Data curation, M.M.; Writing—original draft, A.A.A. and M.M.; Writing—review & editing, S.K.T.; Visualization, M.M.; Supervision, S.K.T.; Project administration, S.K.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research is funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R308).

**Data Availability Statement:** Data are available at (<https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64>, accessed on 25 January 2025).

**Acknowledgments:** Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R308), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Conflicts of Interest:** The authors declare no competing interests.

## References

1. Ahmed Ali, K.; Ahmad, M.I.; Yusup, Y. Issues, Impacts, and Mitigations of Carbon Dioxide Emissions in the Building Sector. *Sustainability* **2020**, *12*, 7427. [CrossRef]
2. Chandio, A.A.; Akram, W.; Ahmad, F.; Ahmad, M. Dynamic relationship among agriculture-energy-forestry and carbon dioxide (CO<sub>2</sub>) emissions: Empirical evidence from China. *Environ. Sci. Pollut. Res.* **2020**, *27*, 34078–34089. [CrossRef] [PubMed]
3. Payne, J.E. The convergence of carbon dioxide emissions: A survey of the empirical literature. *J. Econ. Stud.* **2020**, *47*, 1757–1785. [CrossRef]
4. Dogru, T.; Bulut, U.; Kocak, E.; Isik, C.; Suess, C.; Sirakaya-Turk, E. The nexus between tourism, economic growth, renewable energy consumption, and carbon dioxide emissions: Contemporary evidence from OECD countries. *Environ. Sci. Pollut. Res.* **2020**, *27*, 40930–40948. [CrossRef]
5. Osobajo, O.A.; Otitoju, A.; Otitoju, M.A.; Oke, A. The Impact of Energy Consumption and Economic Growth on Carbon Dioxide Emissions. *Sustainability* **2020**, *12*, 7965. [CrossRef]
6. Yang, S.; Yang, D.; Shi, W.; Deng, C.; Chen, C.; Feng, S. Global evaluation of carbon neutrality and peak carbon dioxide emissions: Current challenges and future outlook. *Environ. Sci. Pollut. Res.* **2023**, *30*, 81725–81744. [CrossRef]
7. Duan, S.; Liu, Y.; Li, L.; Pan, Y. Prediction of Atmospheric Carbon Dioxide Radiative Transfer Model based on Machine Learning. *Front. Comput. Intell. Syst.* **2023**, *6*, 132–136. [CrossRef]
8. Li, S.; Siu, Y.W.; Zhao, G. Driving Factors of CO<sub>2</sub> Emissions: Further Study Based on Machine Learning. *Front. Environ. Sci.* **2021**, *9*, 721517. [CrossRef]
9. Najmi, M.; Ayari, M.A.; Sadeghsalehi, H.; Vaferi, B.; Khandakar, A.; Chowdhury, M.E.H.; Rahman, T.; Jawhar, Z.H. Estimating the Dissolution of Anticancer Drugs in Supercritical Carbon Dioxide with a Stacked Machine Learning Model. *Pharmaceutics* **2022**, *14*, 1632. [CrossRef]
10. Liu, H.; Shen, L. Forecasting carbon price using empirical wavelet transform and gated recurrent unit neural network. *Carbon Manag.* **2020**, *11*, 25–37. [CrossRef]
11. Guo, J.; Li, J.; Sato, Y.; Yan, Z. A Gated Recurrent Unit Model with Fibonacci Attenuation Particle Swarm Optimization for Carbon Emission Prediction. *Processes* **2024**, *12*, 1063. [CrossRef]
12. Qin, X.; Hu, X.; Liu, H.; Shi, W.; Cui, J. A Combined Gated Recurrent Unit and Multi-Layer Perception Neural Network Model for Predicting Shale Gas Production. *Processes* **2023**, *11*, 806. [CrossRef]
13. Yun, P.; Zhang, C.; Wu, Y.; Yang, Y. Forecasting Carbon Dioxide Price Using a Time-Varying High-Order Moment Hybrid Model of NAGARCHSK and Gated Recurrent Unit Network. *Int. J. Environ. Res. Public Health* **2022**, *19*, 899. [CrossRef] [PubMed]
14. Alhussan, A.; Towfek, S. 5G Resource Allocation Using Feature Selection and Greylag Goose Optimization Algorithm. *CMC* **2024**, *80*, 1179–1201. [CrossRef]
15. Hussain, K.; Salleh, M.N.M.; Cheng, S.; Shi, Y. On the exploration and exploitation in popular swarm-based metaheuristic algorithms. *Neural Comput. Applic.* **2019**, *31*, 7665–7683. [CrossRef]
16. Xu, J.; Zhang, J. Exploration-exploitation tradeoffs in metaheuristics: Survey and analysis. In Proceedings of the 33rd Chinese Control Conference, Nanjing, China, 28–30 July 2014; pp. 8633–8638. [CrossRef]
17. Saeed, M.A.; Ibrahim, A.; El-Kenawy, E.-S.M.; Abdelhamid, A.A.; El-Said, M.; Abualigah, L.; Alharbi, A.H.; Khafaga, D.S.; Elbaksawi, O. Forecasting wind power based on an improved al-Biruni Earth radius metaheuristic optimization algorithm. *Front. Energy Res.* **2023**, *11*. [CrossRef]
18. Alharbi, A.H. Classification of monkeypox images using Al-Biruni earth radius optimization with deep convolutional neural network. *AIP Adv.* **2024**, *14*, 065133. [CrossRef]

19. Turja, A.I.; Khan, I.A.; Rahman, S.; Mustakim, A.; Hossain, M.I.; Ehsan, M.M.; Khan, Y. Machine learning-based multi-objective optimization and thermal assessment of supercritical CO<sub>2</sub> Rankine cycles for gas turbine waste heat recovery. *Energy AI* **2024**, *16*, 100372. [CrossRef]
20. Li, X.; Wu, Y.; Wei, M.; Guo, Y.; Yu, Z.; Wang, H.; Li, Z.; Fan, H. A novel index of functional connectivity: Phase lag based on Wilcoxon signed rank test. *Cogn. Neurodyn.* **2021**, *15*, 621–636. [CrossRef]
21. Martín, S.; Quintana, B.; Barrientos, D. Wilcoxon signed-rank-based technique for the pulse-shape analysis of HPGe detectors. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* **2016**, *823*, 32–40. [CrossRef]
22. Saheed, Y.K.; Balogun, B.F.; Odunayo, B.J.; Abdulsalam, M. Microarray Gene Expression Data Classification via Wilcoxon Sign Rank Sum and Novel Grey Wolf Optimized Ensemble Learning Models. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **2023**, *20*, 3575–3587. [CrossRef]
23. Bertinetto, C.; Engel, J.; Jansen, J. ANOVA simultaneous component analysis: A tutorial review. *Anal. Chim. Acta X* **2020**, *6*, 100061. [CrossRef]
24. Liu, Q.; Wang, L. t-Test and ANOVA for data with ceiling and/or floor effects. *Behav. Res.* **2021**, *53*, 264–277. [CrossRef] [PubMed]
25. Muhammed, H.Z.; Almetwally, E.M. Bayesian and Non-Bayesian Estimation for the Shape Parameters of New Versions of Bivariate Inverse Weibull Distribution based on Progressive Type II Censoring. *Comput. J. Math. Stat. Sci.* **2024**, *3*, 85–111. [CrossRef]
26. Yao, L.; Zhang, Z.; Li, Y.; Zhuo, J.; Chen, Z.; Lin, Z.; Liu, H.; Yao, Z. Precise prediction of CO<sub>2</sub> separation performance of metal-organic framework mixed matrix membranes based on feature selection and machine learning. *Sep. Purif. Technol.* **2024**, *349*, 127894. [CrossRef]
27. Zhang, X.; Sun, J.; Zhang, X.; Wang, F. Assessment and regression of carbon emissions from the building and construction sector in China: A provincial study using machine learning. *J. Clean. Prod.* **2024**, *450*, 141903. [CrossRef]
28. Su, S.; Zang, Z.; Yuan, J.; Pan, X.; Shan, M. Considering critical building materials for embodied carbon emissions in buildings: A machine learning-based prediction model and tool. *Case Stud. Constr. Mater.* **2024**, *20*, e02887. [CrossRef]
29. Qiao, Q.; Eskandari, H.; Saadatmand, H.; Sahraei, M.A. An interpretable multi-stage forecasting framework for energy consumption and CO<sub>2</sub> emissions for the transportation sector. *Energy* **2024**, *286*, 129499. [CrossRef]
30. Moon, S.; Lee, J.; Kim, H.J.; Kim, J.H.; Park, S. Study on CO<sub>2</sub> Emission Assessment of Heavy-Duty and Ultra-Heavy-Duty Vehicles Using Machine Learning. *Int. J. Automot. Technol.* **2024**, *25*, 651–661. [CrossRef]
31. Luo, H.; Wang, C.; Li, C.; Meng, X.; Yang, X.; Tan, Q. Multi-scale carbon emission characterization and prediction based on land use and interpretable machine learning model: A case study of the Yangtze River Delta Region, China. *Appl. Energy* **2024**, *360*, 122819. [CrossRef]
32. Giannelos, S.; Bellizio, F.; Strbac, G.; Zhang, T. Machine learning approaches for predictions of CO<sub>2</sub> emissions in the building sector. *Electr. Power Syst. Res.* **2024**, *235*, 110735. [CrossRef]
33. Chen, Y.; Xie, Y.; Dang, X.; Huang, B.; Wu, C.; Jiao, D. Spatiotemporal prediction of carbon emissions using a hybrid deep learning model considering temporal and spatial correlations. *Environ. Model. Softw.* **2024**, *172*, 105937. [CrossRef]
34. Koca Akkaya, E.; Akkaya, A.V. Development and performance comparison of optimized machine learning-based regression models for predicting energy-related carbon dioxide emissions. *Environ. Sci. Pollut. Res.* **2023**, *30*, 122381–122392. [CrossRef]
35. Kumari, S.; Singh, S.K. Machine learning-based time series models for effective CO<sub>2</sub> emission prediction in India. *Environ. Sci. Pollut. Res.* **2023**, *30*, 116601–116616. [CrossRef]
36. Ağbulut, Ü. Forecasting of transportation-related energy demand and CO<sub>2</sub> emissions in Turkey with different machine learning algorithms. *Sustain. Prod. Consum.* **2022**, *29*, 141–157. [CrossRef]
37. Moradi, M.H.; Brutsche, M.; Wenig, M.; Wagner, U.; Koch, T. Marine route optimization using reinforcement learning approach to reduce fuel consumption and consequently minimize CO<sub>2</sub> emissions. *Ocean Eng.* **2022**, *259*, 111882. [CrossRef]
38. Bakay, M.S.; Ağbulut, Ü. Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. *J. Clean. Prod.* **2021**, *285*, 125324. [CrossRef]
39. Alfaseeh, L.; Tu, R.; Farooq, B.; Hatzopoulou, M. Greenhouse gas emission prediction on road network using deep sequence learning. *Transp. Res. Part D Transp. Environ.* **2020**, *88*, 102593. [CrossRef]
40. Sun, W.; Wang, C.; Zhang, C. Factor analysis and forecasting of CO<sub>2</sub> emissions in Hebei, using extreme learning machine based on particle swarm optimization. *J. Clean. Prod.* **2017**, *162*, 1095–1101. [CrossRef]
41. Open Government Portal, C.N.R. Fuel Consumption Ratings-Open Government Portal. 2024. Available online: <https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64> (accessed on 25 January 2025).
42. Yesbolova, A.Y.; Abdulova, T.; Nurgabylov, M.; Yessenbekova, S.; Turalina, S.; Baytaeva, G.; Myrzabekkyzy, K. Analysis of the Effect of Renewable Energy Consumption and Industrial Production on CO<sub>2</sub> Emissions in Turkic Republics by Panel Data Analysis Method. *Int. J. Energy Econ. Policy* **2024**, *14*, 480–487. [CrossRef]

43. Ying, Z.; Qiu, Q.; Ye, J.; Chen, H.; Zhao, J.; Shen, Y.; Chu, B.; Gao, H.; Zhang, S. Mechanism, performance enhancement, and economic feasibility of CO<sub>2</sub> microbial electrosynthesis systems: A data-driven analysis of research topics and trends. *Renew. Sustain. Energy Rev.* **2024**, *202*, 114704. [[CrossRef](#)]
44. Yao, X.; Zhang, H.; Wang, X.; Jiang, Y.; Zhang, Y.; Na, X. Which model is more efficient in carbon emission prediction research? A comparative study of deep learning models, machine learning models, and econometric models. *Environ. Sci. Pollut. Res.* **2024**, *31*, 19500–19515. [[CrossRef](#)]
45. Sattari, M.T.; Apaydin, H.; Shamshirband, S. Performance Evaluation of Deep Learning-Based Gated Recurrent Units (GRUs) and Tree-Based Models for Estimating ETo by Using Limited Meteorological Variables. *Mathematics* **2020**, *8*, 972. [[CrossRef](#)]
46. Rezaei, M.; Mohammadifar, A.; Gholami, H.; Mina, M.; Riksen, M.J.P.M.; Ritsema, C. Mapping of the wind erodible fraction of soil by bidirectional gated recurrent unit (BiGRU) and bidirectional recurrent neural network (BiRNN) deep learning models. *Catena* **2023**, *223*, 106953. [[CrossRef](#)]
47. Zeng, C.; Ma, C.; Wang, K.; Cui, Z. Parking Occupancy Prediction Method Based on Multi Factors and Stacked GRU-LSTM. *IEEE Access* **2022**, *10*, 47361–47370. [[CrossRef](#)]
48. Chen, J.; Zhang, J.; Chen, H.; Zhao, Y.; Wang, H. A TDV attention-based BiGRU network for AIS-based vessel trajectory prediction. *iScience* **2023**, *26*, 106383. [[CrossRef](#)]
49. Eslam, A.; Abdelfattah, M.G.; El-Kenawy, E.-S.M.; Moustafa, H.E.-D. Optimization of Feature Selection Using Greylag Goose Optimization Algorithm for Monkeypox. *J. Artif. Intell. Eng. Pract.* **2024**, *1*, 1–16. [[CrossRef](#)]
50. Saber, M.; Abdelhamid, A.A.; Ibrahim, A. Metaheuristic Optimization Review: Algorithms and Applications. *J. Artif. Intell. Metaheuristics* **2023**, *3*, 21–30. [[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.