



Article

Predicting CO₂ Emissions from Traffic Vehicles for Sustainable and Smart Environment Using a Deep Learning Model

Abdullah H. Al-Nefaie and Theyazn H. H. Aldhyani



Article

Predicting CO₂ Emissions from Traffic Vehicles for Sustainable and Smart Environment Using a Deep Learning Model

Abdullah H. Al-Nefaie ¹ and Theyazn H. H. Aldhyani ^{2,*} 

¹ Department of Quantitative Methods, School of Business, King Faisal University, Al-Ahsa 31982, Saudi Arabia; aalnefaie@kfu.edu.sa

² Applied College in Abqaiq, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia

* Correspondence: taldhyani@kfu.edu.sa

Abstract: Burning fossil fuels results in emissions of carbon dioxide (CO₂), which significantly contributes to atmospheric changes and climate disturbances. Consequently, people are becoming concerned about the state of the environment, and governments are required to produce precise projections to develop efficient preventive measures. This study makes a significant contribution to the area by modeling and predicting the CO₂ emissions of vehicles using advanced artificial intelligence. The model was constructed using the CO₂ emission by vehicles dataset from Kaggle, which includes different parameters, namely, vehicle class, engine size (L), cylinder transmission, fuel type, fuel consumption city (L/100 km), fuel consumption hwy (L/100 km), fuel consumption comb (L/100 km), fuel consumption comb (mpg), and CO₂ emissions (g/km). To forecast the CO₂ emissions produced by vehicles, a deep learning long short-term memory network (LSTM) model and a bidirectional LSTM (BiLSTM) model were developed. Both models are efficient. Throughout the course of the investigation, the researchers employed four statistical assessment metrics: the mean square error (MSE), the root MSE (RMSE), Pearson's correlation coefficient (R%), and the determination coefficient (R²). Based on the datasets of experiments carried out by Kaggle, the LSTM and BiLSTM models were created and implemented. The data were arbitrarily split into two phases: training, which included 80% of the total data, and testing, which comprised 20% of the total data. The BiLSTM model performed best in terms of accuracy and achieved high prediction values for MSE and RMSE. The BiLSTM model has the greatest attainable (R² = 93.78). In addition, R% was used to locate a connection between the dataset's characteristics to ascertain which characteristics had the highest level of association with CO₂ emissions. An original strategy for the accurate forecasting of carbon emissions was developed as a result of this work. Consequently, policymakers may use this work as a potentially beneficial decision-support tool to create and execute successful environmental policies.

Keywords: artificial intelligence; deep learning; CO₂ emissions; vehicles; environment



Citation: Al-Nefaie, A.H.; Aldhyani, T.H.H. Predicting CO₂ Emissions from Traffic Vehicles for Sustainable and Smart Environment Using a Deep Learning Model. *Sustainability* **2023**, *15*, 7615. <https://doi.org/10.3390/su15097615>

Academic Editor: Yagang Zhang

Received: 27 March 2023

Revised: 1 May 2023

Accepted: 2 May 2023

Published: 5 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Increasing greenhouse gas (GHG) emissions over the last several decades have sparked rising concern about global warming. The energy [1], transportation [2,3], agricultural [4], building [5], waste management [6], and afforestation and reforestation [7] sectors are only some of the human-caused sources of these emissions. According to the International Energy Agency [8], transportation is the second largest sector responsible for emitting GHGs. Typically reported as carbon dioxide equivalent (CO₂e), nitrogen oxide (N₂O), methane (CH₄), and carbon dioxide (CO₂) are the principal GHG emissions from automobiles [9]. Globally, transportation activities produce 75% of total environmental CO₂ emissions [10]. Consequently, the reduction of emissions from transportation is a worldwide objective against climate change [11]. Assessing urban highways and exposure to pollutant emissions from motor traffic is important when considering economic, social, and environmental

goals, especially in emerging countries [12]. Figure 1 shows how CO₂ is emitted from vehicles.

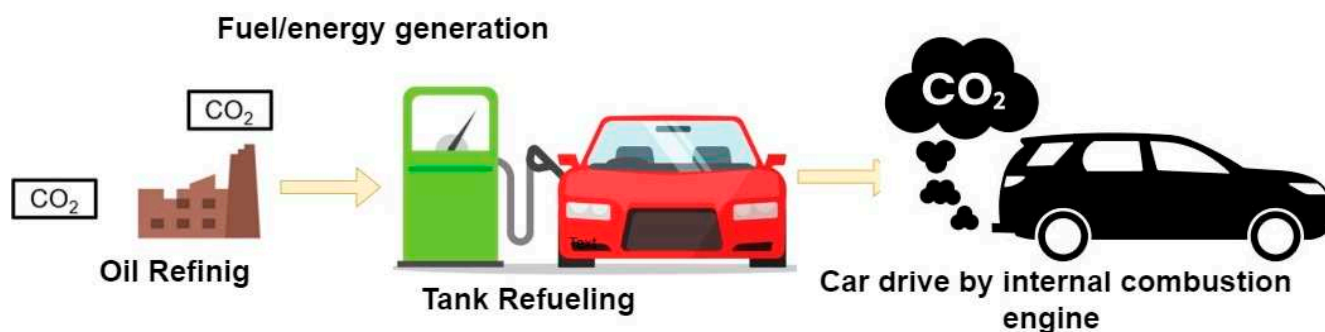


Figure 1. Carbon dioxide from vehicles.

Building transportation infrastructures, such as expressways and motorways, that connect cities is essential to any expanding economy. Congestion on the nation's highways is a direct consequence of the rapid expansion of infrastructure built to accommodate the development of available modes of transportation. Due to this, a discernible decline in the quality of the air around motorways, intersections, and tollways has been observed. Vehicle exhausts are the primary contributor to traffic emissions, such as carbon monoxide (CO), which are a substantial factor in the overall amount of air pollution caused by these infrastructures. Spatial prediction models are an outstanding decision-making assistance tool because of their ability to estimate and simulate the effects of traffic emissions on road networks [13–15]. An excessive amount of traffic may have various undesirable impacts [16,17], including an increase in both noise and gaseous pollution. Humans exposed to high amounts of CO₂ have a significantly increased chance of contracting various diseases and conditions [18–20], such as cancer, heart disease, respiratory problems, and preterm deliveries.

Given the rapid development of computer technology in recent years, a significant number of studies have focused on the use of machine learning and deep learning to forecast exhaust emissions [21] and have used regression analysis to predict CO₂ emissions of light-duty diesel trucks (LDDTs); the correlation coefficient (R%) between the regression-equation-based CO₂ projections and the actual CO₂ measurements was 0.93%. Maksymilian et al. [22] collected RDE data from hybrid electric automobiles, tested the accuracy of several machine learning models to estimate CO₂ emissions, and determined that the most effective approach is the Gaussian process regression. In addition, CO₂ emissions produced by hybrid electric automobiles have not always increased in proportion to the vehicle's rate of acceleration. In the past several years, artificial neural networks (ANNs) have been widely applied in the literature as a method for estimating numbers in real applications with a high degree of accuracy [23]. Hashemi and Clark [24] trained an ANN model to effectively predict CO₂ emissions from heavy-duty diesel automobiles using the properties of axle speed, torque, and their derivatives. Jigu et al. [25] combined an ANN model and a vehicle dynamics model to anticipate the short-term CO₂ emissions of light-duty diesel cars. The CO₂ calculation is based only on these two parameters—engine speed and torque—instead of considering the impact of actual driving conditions. In some studies [26–28], ANN has been used to predict the pollutants released when a vehicle's exhaust is burned. The problem with these studies is that the models used to create forecasts only considered the current status of the car's engine and the condition of the road. Furthermore, automobile emissions have a high temporal dependence. However, even after time alignment, time error still occurs in the measurement between the value of the exhaust gas and the value of the engine while it operates. Given this factor, we could not depend on these studies for prediction by simply using currently accessible information.

Several models may be used to anticipate CO₂ emissions from transportation, and the literature [29–31] discusses the many methods that can be used.

Early traffic CO₂ modeling is established based on traditional approaches that are mostly dependent on data sampling and technologies, such as GPS. The simulation of traffic emission distribution in a region may be achieved by combining different themed maps and vehicle emission equations, resulting in the generation of informative maps that can aid in making informed judgments [29]. For land-use regression analysis, recently developed methods largely depend on statistical and soft computing algorithms [32]. These computational and statistical approaches may use various factors pertaining to traffic and road layout as inputs. Given that an overwhelming majority of these models are constructed using experimental data, they are highly sensitive to alterations in traffic flow, measurement techniques, and geographical locations [32]. These models cannot be generalized since they depend on the local environment, which considers factors, such as the kind and make of the vehicle and the weather [33,34]. A method that uses surveyed location information acquired from automobiles as its data source for detecting road geometry components has been established [35].

Long short-term memory networks (LSTMs) have been established by academics to solve gradient disappearing and gradient clipping [36]. Because of this, they are far better than people at estimating the amount of CO₂ released by cars. Tao predicted the instantaneous CO₂ emissions of taxis using an LSTM-based vehicle emission model with more precision than other state-of-the-art techniques [37]. Yang et al. [38] found that LSTM performs better than other models in forecasting short-term fluctuations in NO_x emissions. Estimating CO₂ emissions from light-duty diesel trucks (LDDTs) using sequence models, on the other hand, is almost rarely encountered in the published literature.

In [39], researchers utilized a GBR model to calculate the emissions and fuel utilization of a hypothetical high-speed vehicle. Overall, we're happy with the outcomes. Nevertheless, the prediction model only included information about when the internal combustion engine vehicles (ICE) were turned on, which contradicts our findings. Hence, the authors did not account for changes in the status of the power train that occur at HVs.

Two light-duty diesel-powered autos' CO₂ emissions were predicted using LSTM, DNN, and CNN in a comparative study [40]. The data set was retrieved from the car's onboard diagnostics (OBD2) connector. Of the tested models, the LSTM model earned the lowest RMSE score (9.30). The authors also investigated the impact of noisy data in the input data, which may reduce the model's accuracy by as much as 30%.

Although traditional micro and macro estimate algorithms are employed elsewhere, Ref. [41] used a number of machine learning techniques to anticipate the CO₂ emission rate of hybrid vehicles. Compared to the other models, the Gaussian process regression (GPR) model achieved the best accuracy at 69%. The authors conclude that there is an issue with the current crop of emission models since they do not provide consistent outcomes. Hybrid cars' inconsistent CO₂ emissions are to blame for this issue. The results of this research provide a significant contribution to the search for a remedy to this issue.

Emission prediction models for several vehicle pollutants may be built using PEMS, and the authors of research [42] suggest using a parallel attention-based LSTM (or PA-LSTM). The training time was reduced since the model's convergence speed was increased by using the two-layer attention spatial encoding approach. The results demonstrated that the PA-LSTM outperformed every other ML mode, with an accuracy of 94.6% being attained.

Many ML models were used to calculate the CO₂ output of a diesel-powered vehicle [43]. Construction equipment chassis movement has been shown to correlate significantly with emission rates, which the authors explored. The results indicated that RF was the top-performing algorithm, with a 94% success rate.

LSTM models have been used to anticipate a variety of real-time applications, such as [44], which used LSTM models to predict air pollution, and [45], which used LSTM models to predict the performance of thermal power plants. Ref. [46] forecast carbon

emissions, Ref. [47] forecast future exports, Ref. [48] analyze marketing data, Ref. [49] forecast electricity load, and Ref. [50] forecast stock prices.

One of the primary contributors to the difficulties caused by pollution in the atmosphere is the ever-increasing number of cars. A long- and short-term forecast of the exhaust emission from on-road vehicles that is both timely and accurate might help prevent air pollution, the protection of public health, and the decision-making process that governments use for environmental management. Because of the inherent unpredictability and imbalance that comes with the nature of meteorological elements and traffic flow, the emission of vehicle exhaust has significant non-stationary and nonlinear characteristics. Hence, accurate long- and short-term vehicle exhaust emission prediction has a number of obstacles, including the long- and short-term temporal dependencies and intricate nonlinear connection on various emission gases, including carbon monoxide (CO). We propose a unique hybrid deep learning framework, namely LSTM and BiLSTM models, to efficiently forecast long- and short-term multivariate vehicle exhaust emissions. This will allow us to overcome the hard challenges that have been presented.

The purpose of this research is to build a model using artificial intelligence technology that is capable of predicting CO₂ emissions from automobiles, with an added focus on Saudi Arabia, which is one of the top oil-producing nations in the world. According to the numbers that have been made public, there are now more than 12 million automobiles on the land that comprises the Kingdom of Saudi Arabia. This number is expected to rise to 25 million automobiles by the year 2030. The creation of such a model may contribute to the protection of the environment and the limitation of the spread of CO₂. The main contributions of the research article are as follows:

1. In general, this effort contributes to the achievement of a few of the Sustainable Development Goals set out by the United Nations. It is directly associated with Goal No. 13, which is titled “climate action,” and it is also assisting in an indirect way to attain other objectives.
2. More specifically, in this study, a highly effective deep learning long short-term memory network (LSTM) model and a bidirectional LSTM (BiLSTM) model have been constructed to forecast CO₂ emissions from traffic cars.
3. Rough k-means clustering is proposed as a preprocessing approach to handle the outliers in the entire dataset for improving the deep learning models.
4. These models also estimate the impact of the population of cars on CO₂ production. As a novel method, sensitivity analysis was added to the model that was constructed in order to investigate the typical CO₂ emissions produced by traffic vehicles.
5. The performance of the present model in estimating CO₂ emissions was found to be superior to that of the other available models.

2. Materials and Methods

The framework for predicting and modeling CO₂ emissions from vehicle traffic is displayed in Figure 2. The step-by-step modeling technique is presented in its entirety in the proposed framework.

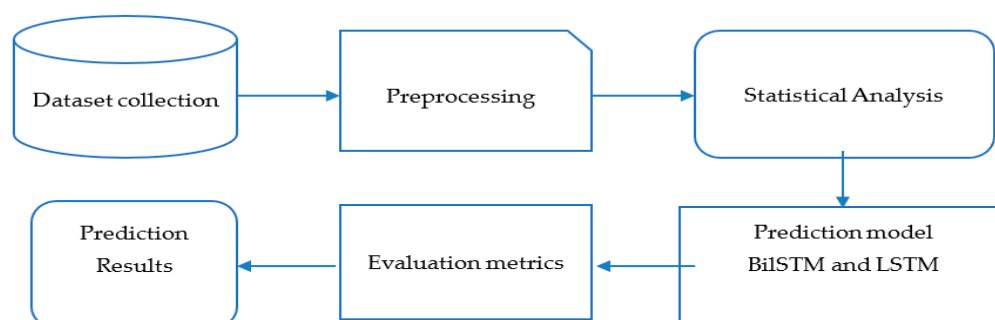


Figure 2. Framework of carbon dioxide emission prediction.

2.1. Dataset

This dataset contains the official record of the CO₂ emission data of various cars with different features. This data collection accurately depicts the range of possibilities for a vehicle's CO₂ emissions as a function of its many attributes. The information was obtained from the Canadian government's open data portal. We prepared the latest version for convenience. The information provided spans 7 years. A total of 12 columns and 7385 rows were provided. Several acronyms were utilized to shorten the list of characteristics. Table 1 shows the features of the dataset, where float64 is consisting of both positive and negative values separated by a decimal point and int64 is the real number.

Table 1. Features of the dataset.

#Variable	Types
#Model	object
#Vehicle Class	Object
#Engine Size (dm ³)	float64
#Cylinders	int64
#Transmission	int64
#Fuel Type	int64
#Fuel Consumption City (L/100 km)	float64
#Fuel Consumption Hwy (L/100 km)	float64
#Fuel Consumption Comb (L/100 km)	float64
#Fuel Consumption Comb (mpg)	float64
#CO ₂ Emissions (g/km)	int64

Vehicles are classified according to size. Figure 3 illustrates the vehicle class type. A total of 16 separate vehicle classes were established. We classified them as hatchbacks, sedans, SUVs, or trucks. The plot shows that the larger the cars, the more CO₂ they emit. A sedan is a four-door automobile manufactured in a three-box design and has a boot that is made separately, whereas a hatchback is a four-door automobile produced in a two-box design. Several minor distinctions are made between a hatchback and a sedan when comparing characteristics, such as size, comfort, and fuel efficiency, among others. Sports utility vehicles are commonly referred to as SUVs. An SUV is a kind of motor vehicle that combines the passenger capacity of a minivan with the off-road prowess and towing capacity of a pickup truck. Given these factors, SUVs are fantastic choices for great outdoor adventures. Notably, the larger the automobiles, the more carbon dioxide they produce. The lowest and higher quartiles of CO₂ emissions are represented by the bottom and upper sides of the box, respectively.

The median and mean levels of CO₂ emissions for each vehicle are shown by the line that divides the box in half horizontally. The lowest and highest quartile CO₂ emissions for each vehicle are shown by the two lines that extend beyond the box. The areas where the dots cross the maximum quartile represent areas with CO₂ emissions that are higher than the maximum.

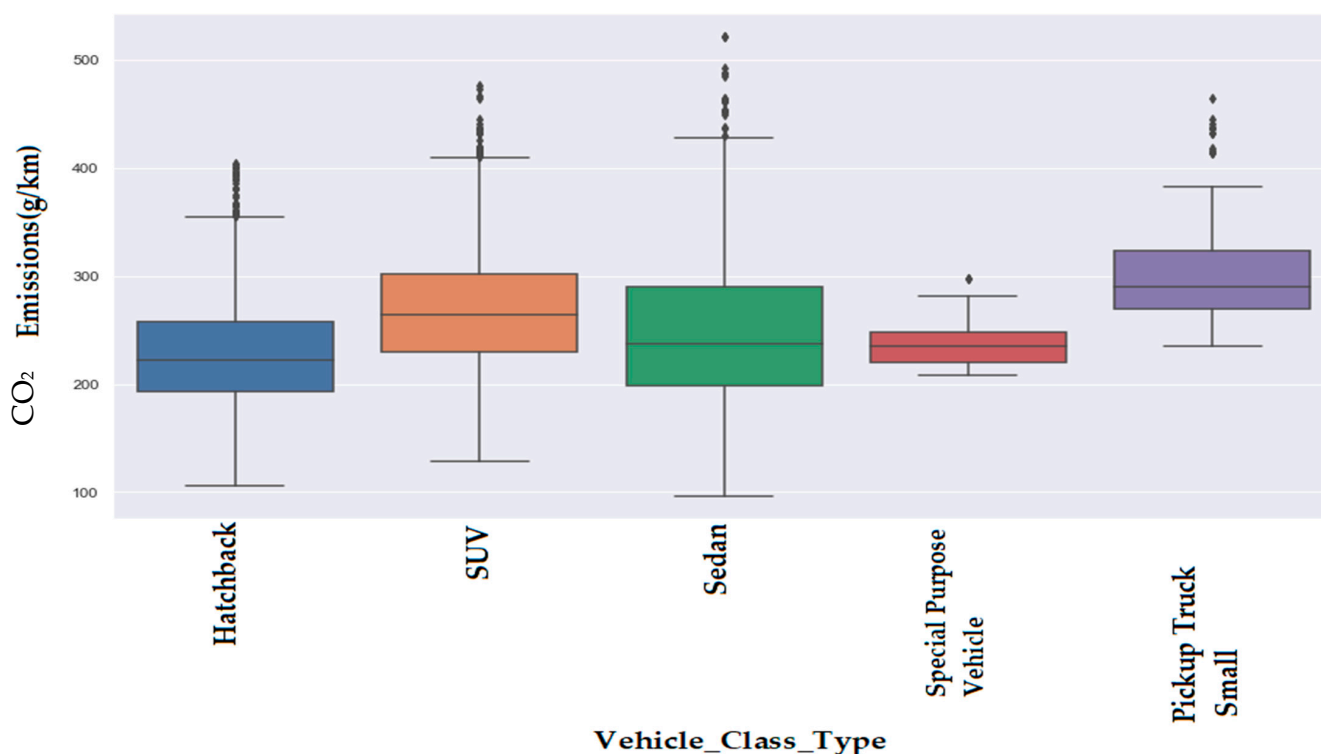


Figure 3. Relationship between vehicle classes and carbon dioxide emissions.

2.2. Preprocessing

The term “data preparation” refers to the actions that must be performed to convert or encode data so that it can be read and understood by a computer. To produce accurate and precise model predictions, the algorithm that underpins it must have the capacity to swiftly analyze the qualities of the data. Given that the quality of the data and useful information derived from it has a direct influence on our model’s potential to learn, the preprocessing of our data before feeding it into our model is key. Data preprocessing is one of the most important stages of machine learning. The phase in the process of data preparation, which includes filling in missing values, smoothing noisy data, resolving inconsistencies, and eliminating outliers, is called “filling in missing values”. A total of 6281 entries and 12 features are now provided in the dataset, and neither missing nor duplicate values are present. Data preparation involves combining data obtained from various sources into a single, more extensive data storage facility, such as a data warehouse. The data processing steps are presented in Figure 4.



Figure 4. Preprocessing steps.

Normalizing a group of independent variables or data components may be accomplished using a method called feature scaling. In the field of data processing, this phase, which is often performed during the data preparation stage, is also known as data nor-

malization. The simplest technique, which is also known as min-max scaling or min-max normalization, involves rescaling the range of features to scale the range in [0, 1]. Other names for this method are min-max scaling and min-max normalization. The formula for universal normalization is as follows:

$$\frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

2.3. Rough K-Means Clustering (RKM)

A straightforward k-means clustering serves as the foundation for the RKM clustering methodology that has been presented [51]. Joshi et al. [52] improved the algorithm of initial concept by computing an approximate centroid by utilizing ratios of distances as the new recommended method to discriminate between distances that were otherwise comparable to one another in order to manage high dimensional data. In their study on ambiguous items related to intrusion detection, Ref. [53] employed the rough k-means and ECM clustering techniques. The rough k-means method is a methodology that was developed with the intention of identifying the unclear items that belong to the top limit of clusters.

Create a lower approximation cluster and a higher approximation cluster using the data. Each variable is represented by its approximate k-mean.

- (P1) An object $\vec{x}$ can be part of a lower approximation of at most one cluster.
 (P2) $\vec{x} \in \underline{A}(\vec{c}_i) \Rightarrow \vec{x} \in \overline{A}(\vec{c}_i)$
 (P3) An object $\vec{x}$ is not part of any lower approximation. (2)
- $\Updownarrow$
 $\vec{x}$ belongs to the upper approximation of two or more clusters.

Soft clustering as a concept in general rough means k-test. All objects are given values for w_{lower} and w_{upper} once the algorithm has run its course. Let the distance between each item vector and the cluster center, denoted by $\vec{v}$ let $d(\vec{v}, \vec{c}_j)$ be considered. The $d(\vec{v}, \vec{c}_i) = \min_{1 \leq j \leq k} d(\vec{v}, \vec{c}_j)$, The proportions $d(\vec{v}, \vec{c}_j)/d(\vec{v}, \vec{c}_i)$, $1 \leq j \leq k$, are used to identify the participants in. So, if we set $T = \{j: d(\vec{v}, \vec{c}_j)/d(\vec{v}, \vec{c}_i) \geq \text{threshold and } i \neq j\}$ we obtained threshold.

1. In the event that $T = \emptyset$, $\vec{v} \in A(\vec{c}_j)$ and $\vec{v} \in A(\vec{c}_j)$, $\forall j \in T$ moreover, it does not appear in any more fundamental approximation. The aforementioned condition ensures the fulfillment of property (P3).
2. Alternately, if $T = \emptyset$, $\vec{v} \in A(\vec{c}_j)$. Furthermore, $A()$ is a characteristic of (P2), $\vec{v} \in A(\vec{c}_j)$.

The rough k-means method is stable and reliable for dealing with ambiguity because of its design. The crude k-means algorithm has successfully grouped the items into upper-bound and lower-limit categories. The things that are located in the upper limit are considered to be ambiguous, whilst the objects that are located in the lower bound are considered to be right. The upper limit must not be empty, and the objects that are included inside it may have a connection to one or more higher bounds that are found in the cluster numbers. Figure 5 shows the results of RKM for handling the outlier from observation.

The ambiguous objects in CO₂ emission datasets have reduced the performance of deep learning algorithms. When applying the algorithm to the original data, it is observed that the results are not favorable. From the data, it is investigated that there are ambiguous objects that hinder the classification algorithms. These outlier objects are examined by rough k-means clustering to assist in determining the exact ambiguity. The dataset has been clustered for five corresponding clusters, and the RKM algorithm has clustered the outlier objects into upper approximation and lower approximation. Those objects that belong to upper approximation, which belongs to one or more cluster numbers, are excluded (outliers).


```

Number of iterations: 69
Centroid for cluster 0: 373.32
Lower bound for cluster 0:
13,14,16,18,19,36,46,48,50,52,70,72,73,74,75,112,132,173,175,177,181,182,193,21
,3719,3746,3747,3860,3899,3958,3959,3960,3961,3979,4008,4009,4048,4049,4051,405

Boundary region for cluster 0:
15 17 47 51 118 120 122 124 174 176 178 179 194 210 214 216 220 362 387 389 43
Centroid for cluster 1: 310.31
Lower bound for cluster 1:
43,45,49,53,54,55,63,69,71,119,121,123,125,131,134,168,180,184,191,212,241,244
97,1799,1802,1803,1842,1852,1853,1856,1857,1866,1873,1874,1879,1953,1954,1979,
369,3373,3431,3433,3435,3446,3449,3450,3451,3470,3478,3479,3480,3485,3502,350
4627,4644,4645,4647,4651,4652,4662,4664,4668,4669,4672,4705,4710,4711,4726,47
,5849,5850,5851,5853,5854,5856,5857,5858,5859,5860,5861,5862,5868,5869,5870,5
5,7122,7125,7126,7127,7131,7132,7173,7174,7176,7177,7178,7180,7229,7246,7247,

Boundary region for cluster 1:
15 17 35 44 47 51 57 59 64 104 106 108 110 111 116 117 118 120 122 124 141 174
018 3019 3020 3071 3141 3151 3152 3208 3239 3245 3285 3329 3364 3371 3408 3410
6543 6544 6553 6561 6563 6582 6583 6584 6585 6601 6602 6605 6616 6618 6623 662
Centroid for cluster 2: 185.49

```

Figure 5. Snapshot from results of the RKM.

2.4. Statistical Analysis

The primary focuses of statistical analysis are the collection, process, interpretation, and presentation of data. The first thing that must be done in every statistical analysis is to characterize the target phenomena. The population is based on a known time series, as well as data gathered through observing the process as it occurs at various periods. The significance of employing statistical analysis to determine the properties of the data is not forgotten soon. As a result of statistical research conducted to predict CO₂ levels, we investigated statistical models that are less complicated, use few inputs, and cost less to operate. Table 2 displays the findings of the statistical analysis. According to Figure 6, sports vehicles and luxury automobiles are responsible for a greater amount of CO₂ emissions than premium and general-usage cars.

The characteristics that are most indicative of the current CO₂ data collection act as our main research instrument. Figure 7 shows that the data set in question has a significant number of outliers, which is an essential factor to consider. Given that the mean is the highest and the mode is the lowest of the three measures, CO₂ emissions are positively skewed, suggesting that the great majority of observations fall below the mean value [54].

Table 2. Sensitivity analysis results.

	Engine Size	Fuel_ Consumption_ City	Fuel_ Consumption_ Hwy	Fuel_ Consumption_ Comb	Fuel_ Consumption_ Comb1	CO ₂ _ Emissions
Count	6282.00	6282.00	6282.00	6282.00	6282.00	6282.00
Mean	3.16181	12.61022	9.070583	11.01787	27.411016	25.157752
Std	1.36520	3.553066	2.278884	2.9468	7.245318	59.290426
Min	0.9000	4.20000	4.00000	4.100000	11.00000	96.00000
Max	8.40000	30.600000	20.600000	26.100000	69.00000	522.0000

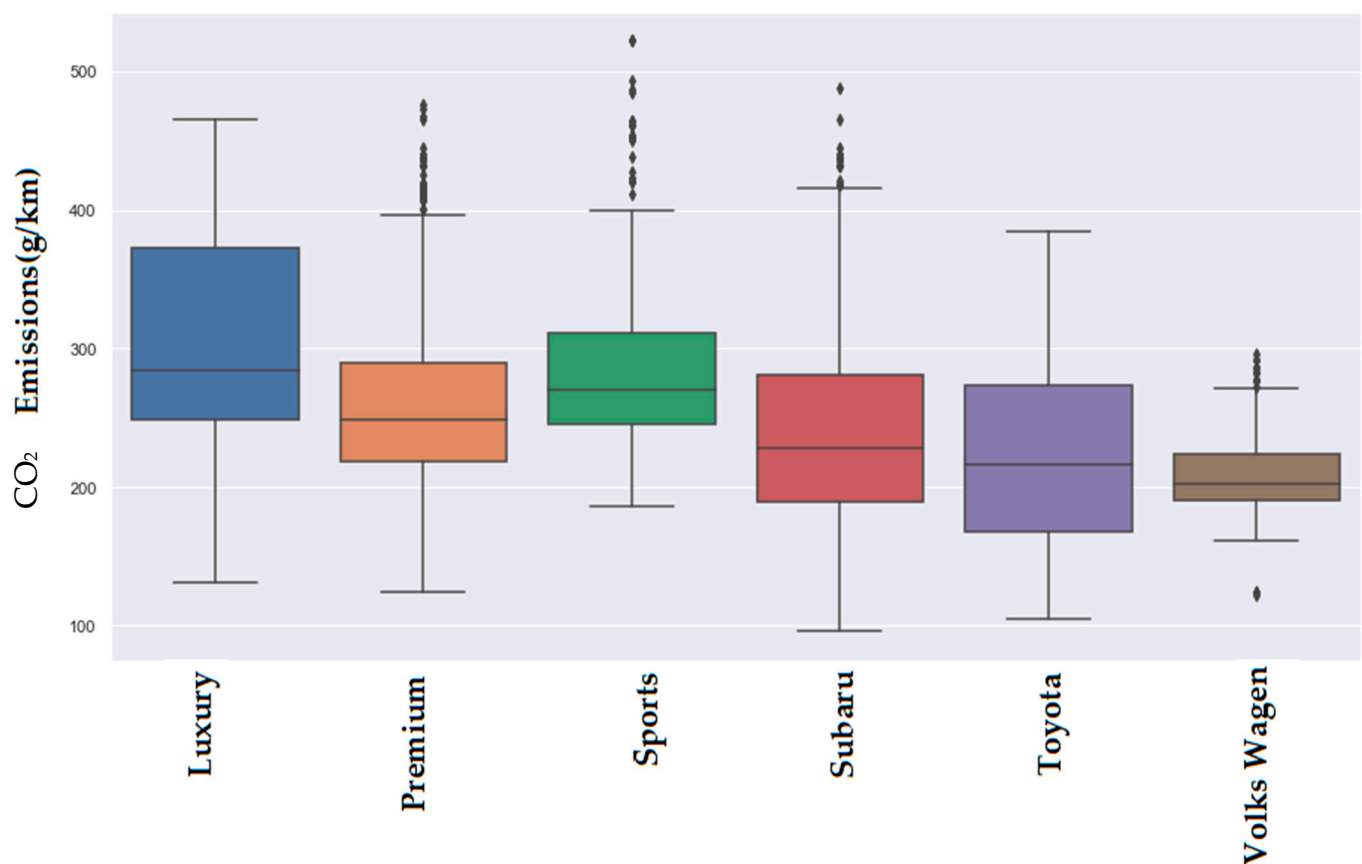


Figure 6. Distribution of the carbon dioxide emissions of class car values from the dataset.

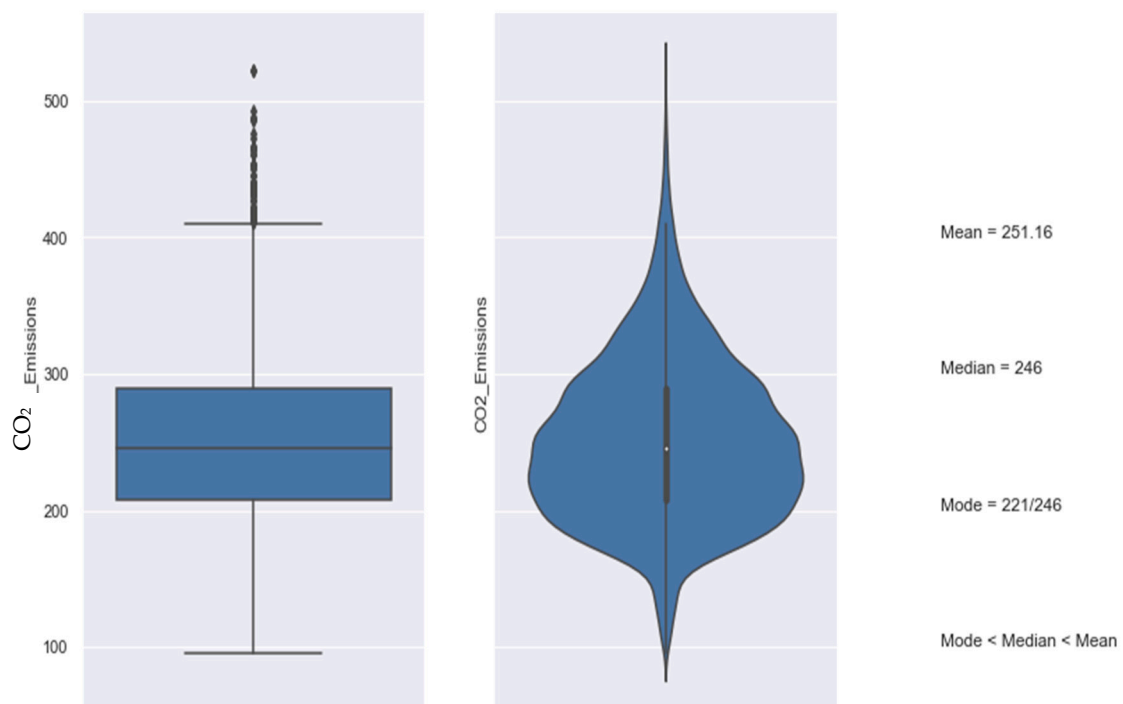


Figure 7. Boxplot of instant carbon dioxide emissions.

Figure 8 displays the average values of the categorical features of the dataset. The most commonly driven automobile model on Canadian roadways is the Ford model. The Ford

F-150 FFV is currently one of the most well-known vehicles on Canadian roads. According to the top ten vehicle class rankings, the SUV-small category is the most popular among Canadian drivers. Top ten transmission: more than one thousand vehicles are equipped with either an AS6 or AS8 type transmission. Types of fuel: most automobiles in Canada operate on fuel types X and Z. CO₂ emissions in relation to make: while most vehicles on Canadian roads are Fords, Bugatti is the brand that produces the highest amount of CO₂ per vehicle. CO₂ emission v/s: the Bugatti Chiron is among the automotive models that release the maximum amount of CO₂ into the atmosphere. CO₂ emissions versus vehicle class: most heavy vehicles, such as vans, SUVs, and pick-up trucks, are among the top emitters of CO₂. Emissions versus transmission: most cars with automatic transmissions emit CO₂. Emissions versus fuel type: vehicles with fuel type E emit the greatest amount of CO₂. CO₂ emissions versus fuel type: vehicles that use fuel type D emit the least amount of CO₂. Where X = regular gasoline; Z = premium gasoline; D = diesel; E = E85; B = electricity.

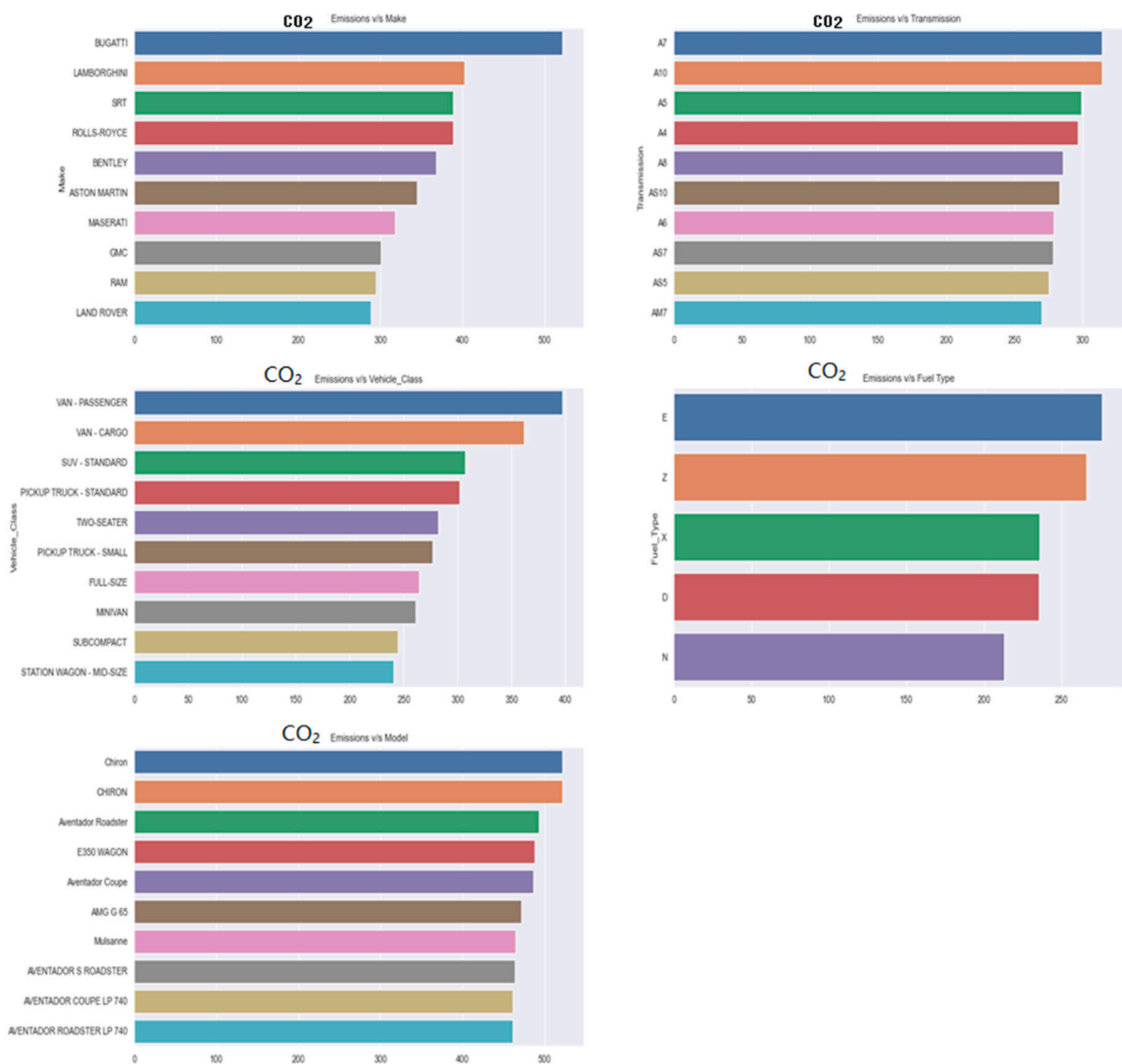


Figure 8. Average of categorical variables and carbon dioxide emissions.

The relationship between the seven investigated factors is shown in Figure 9. The factors with the largest bearing on CO₂ emissions are provided. Using this procedure, we settled on seven separate factors to calculate the CO₂ output. High negative correlations (≤ 0.9) are shown between fuel consumption city, fuel consumption_comb, and CO₂ Emissions, whereas strong negative correlations (≤ 0.8) are observed between cylinder count and engine displacement and CO₂ emissions.

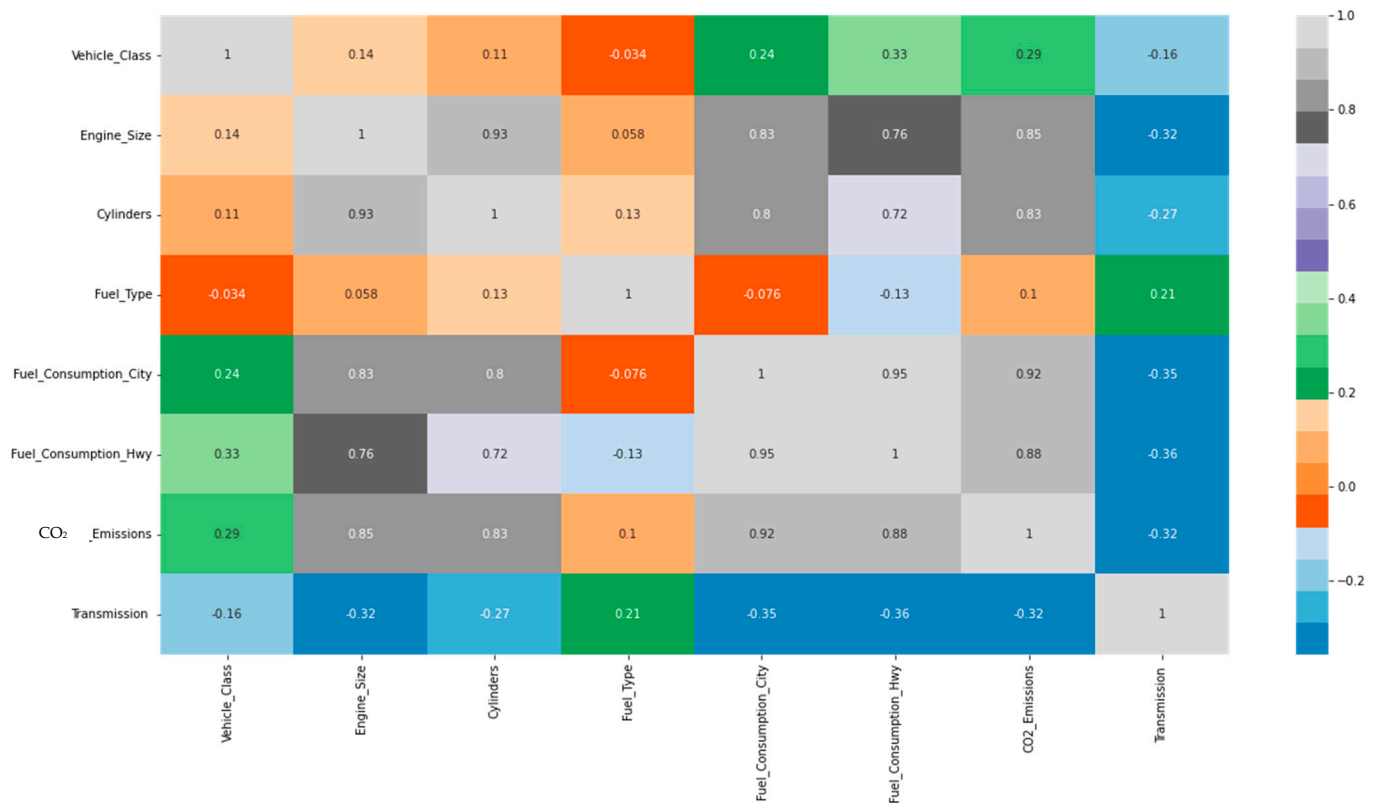


Figure 9. Correlation plot between the features and carbon dioxide emissions.

2.5. Prediction Models

Based on long-series forecasting, RNN is unreliable because of the difficulties posed by gradient explosion and gradient vanishing. Given its superior memory cells, LSTM-type RNN can avoid the problems of gradient explosion and disappearing gradients during the training process. LSTM can control both long- and short-term memories due to its one-of-a-kind hidden state and its three regulating gates. At each time step, LSTM receives, as input, the currently present input value, the previously obtained output value, and the hidden state. The memory state is where any and all information from earlier sequences pertinent to the current state is kept [55,56].

The input gate, the forget gate, and the output gate are shown in Figure 10 shows fundamental LSTM mechanisms, and these gates input the gate i_t , forget gate f_t , and output gate o_t , respectively, which can be expressed using the following mathematical notation:

$$i_t = \sigma(W_i \cdot X_t + W_i \cdot h_{t-1} + b_i) \quad (3)$$

$$f_t = \sigma(W_f \cdot X_t + W_f \cdot h_{t-1} + b_f) \quad (4)$$

$$o_t = \sigma(W_o \cdot X_t + W_o \cdot h_{t-1} + V_o \cdot C_t + b_o) \quad (5)$$

where W_i , W_f , and W_o , each represent weight matrices applied to the combination result of the output value of the previous time step (h_{t-1}), the memory state of the previous

time step (C_t), and the input value of the current time step (X_t), and b_i , b_f , and b_o , each represent the bias applied to that combination result, respectively. To regulate each of these three gates, we employed the sigmoid function as an activation function. This strategy allows the generation of non-negative derivatives that fall between 0 and 1.

$$\bar{c} = \tanh(W_c[h_{t-1}X_t] + b_c) \quad (6)$$

$$C_t = (f_t * C_{t-1} + i_t * \bar{c}) \quad (7)$$

where $\bar{c}$ represents a possible activation function of the tanh type used to change the storage state of memory. If the values of the gates are either exactly zero or extremely close to zero, the gates are closed. If the values fed into the gates do not match the expected values, the gates are opened. If $f_t = 0$ and $i_t > 0$, the memory state of the current input candidate should be forgotten, and instead, the memory state from the previous input candidate should be utilized as the new memory state.

$$h_t = o_t + \tanh(C_t) \quad (8)$$

Moreover, an output gate o_t determines the value read from a given current memory cell h_t . By manipulating the value of the memory state via the movable gate mechanism, the LSTM may extract useful information about the entire series. For this reason, LSTM is often used to predict future electrical loads.

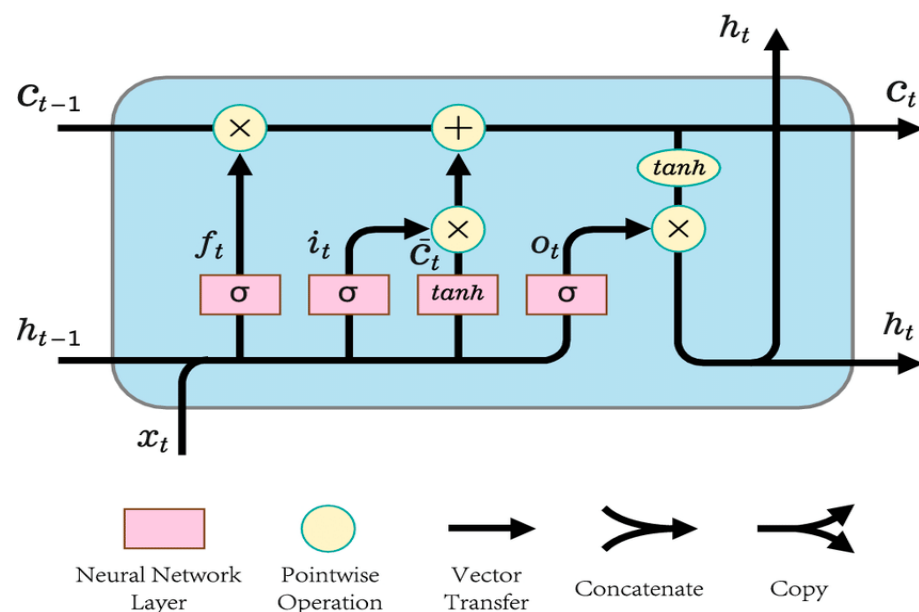


Figure 10. LSTM structure.

Bidirectional LSTM

One variant of LSTM, the BiLSTM, shows potential for both retrospective learning and prospective prediction. BiLSTM incorporates two LSTM layers that propagate in the opposite direction. The first LSTM layer displays series data prospectively, whereas the second displays data retroactively. Each LSTM layer's computation is specified by Equations (3)–(7), as is the case with the LSTM layer shared by all BiLSTM architectures. The second LSTM layer is only inverted in BiLSTM. Figure 11 shows that the two LSTM layers propagate simultaneously in opposite directions. Using a function (f), which may be a sum, concatenation, or average, the BiLSTM combines two hidden state sequences—one from the forward direction and one from the back. BiLSTM develops a new sequence where

each element is a hybrid of past and future information. A diagrammatic representation of the steps taken and the results obtained are as follows:

$$H_t = f(\vec{h}_t : \overleftarrow{h}_t) \quad (9)$$

where $\vec{h}_t$ and $\overleftarrow{h}_t$ represents the forward hidden state at time t , and H represents the backward hidden state at time t . The final series H includes the hidden states from every time step in the past and future. N is the descriptor for the total number of elements in the sequence. In this manner, BiLSTM generates a stream whose elements include both past and future information. It simplifies the processing of series by handling a raw series and can gather all the elements of a series in both directions. Time series forecasting is one area where BiLSTM is superior to LSTM models. In spite of their effectiveness in time series forecasting, LSTM and BiLSTM are restricted in their capacity to learn the complex dynamics of time series due to the discretization of the observation and emission periods in both their components. As a result of the fact that the node is a continuous neural network, an eternally deep network can still be built, which solves the previously addressed problem regarding LSTM. For electricity load forecasting, we conducted an analysis of nodes using both the LSTM and BiLSTM architectures.

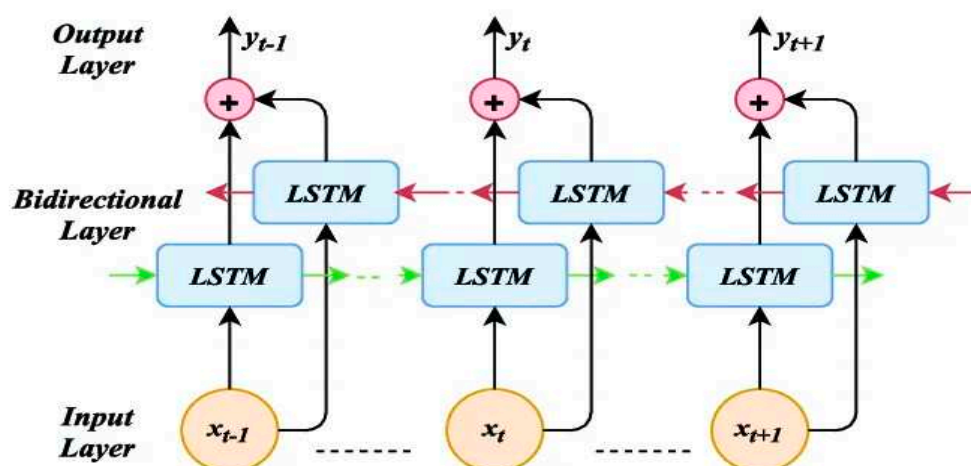


Figure 11. BiLSTM structure.

The predictive capacity of a deep learning model is determined by the hyperparameters and the structure of the model. In this inquiry, we compared various sets of hyperparameters that describe network architectures to assess alternative sub-optimal network models that may be used to simulate vehicular CO₂. Table 3 contains a list of the structures and hyperparameters investigated throughout this study, as well as the domains of the search spaces that correspond to each one.

Table 3. Deep leaning model parameters.

Parameters	Values
Firstlayer	100
Secondlayer	100
Thirdlayer	200
Fourthlayer	200
executionEnvironment	CPU

Table 3. *Cont.*

Parameters	Values
miniBatchSize	20
maxEpochs	100
Optimizer	Adam
Learning rate	0.005

2.6. Evaluation Metrics

Finally, the performance of each model was tested on the test set to evaluate the model's generalization skills in predicting unobserved data. Mean, MSE, RMSE, Pearson's R%, and R2 were used to evaluate the deep learning model. The statistical metrics used to evaluate the forecast models were constructed over the course of this research. These parameters are presented below:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{i,exp} - y_{i,pred})^2 \quad (10)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_{i,exp} - y_{i,pred})^2}{n}} \quad (11)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i,exp} - y_{i,pred})^2}{\sum_{i=1}^n (y_{i,exp} - y_{avg,exp})^2} \quad (12)$$

where $y_{i,exp}$ and $y_{i,pred}$ are the CO₂ observation and prediction values, respectively.

3. Modeling Results

Vehicle traffic and its contaminants are mostly blamed for the unhealthy levels of air pollution, such as CO₂, in populated cities, particularly along toll road corridors. Some models are used to evaluate the consequences of CO₂ emissions from transportation on both humans and the environment, whereas other models are designed to demonstrate the geographical prediction of these emissions. We employed the MATLAB 2020 software for the experiment. The physical configuration of the machine had a 2.8 GHz Intel i7 Quad-Core CPU and 8 GB of memory as part of its components. Compared with previous research, this sample size is still rather small. In contrast, it requires an effective strategy for dealing with overfitting issues. In the initial model, vehicle class, engine size (L), cylinders transmission, fuel type, fuel consumption city (L/100 km), fuel consumption hwy (L/100 km), fuel consumption comb (L/100 km), fuel consumption comb (mpg), and co₂ emissions (g/km) are the input variables utilized for the model. For data normalization, the max-min approach was used. Mean squared error (MSE), root MSE (RMSE), R%, and determination coefficient (R2) are the four statistical characteristics used in the evaluation of the constructed model's capacity in producing accurate predictions. The deep learning models can represent complicated simulations of non-linear issues, such as vehicle CO₂ emission. This capacity was made possible by recent advancements in computing power. The training process of LSTM and biLSTM methods is shown in Figures 12 and 13.

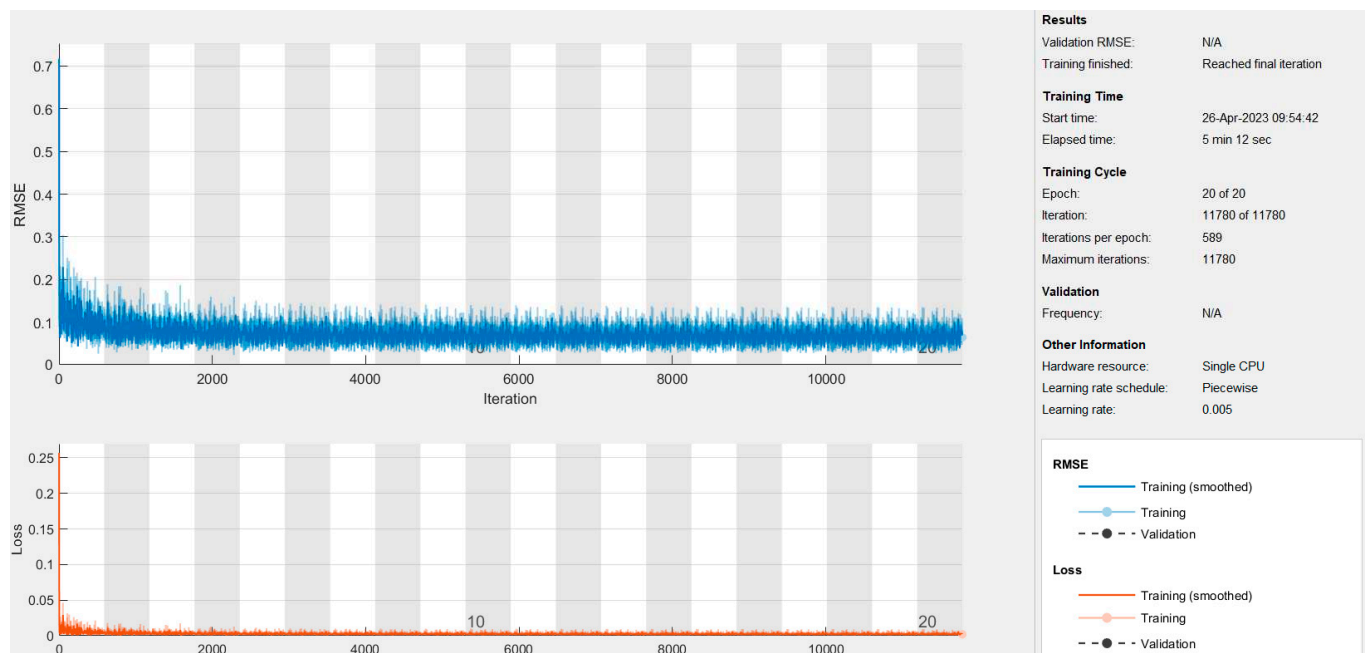


Figure 12. Training process of the long short-term memory network model.

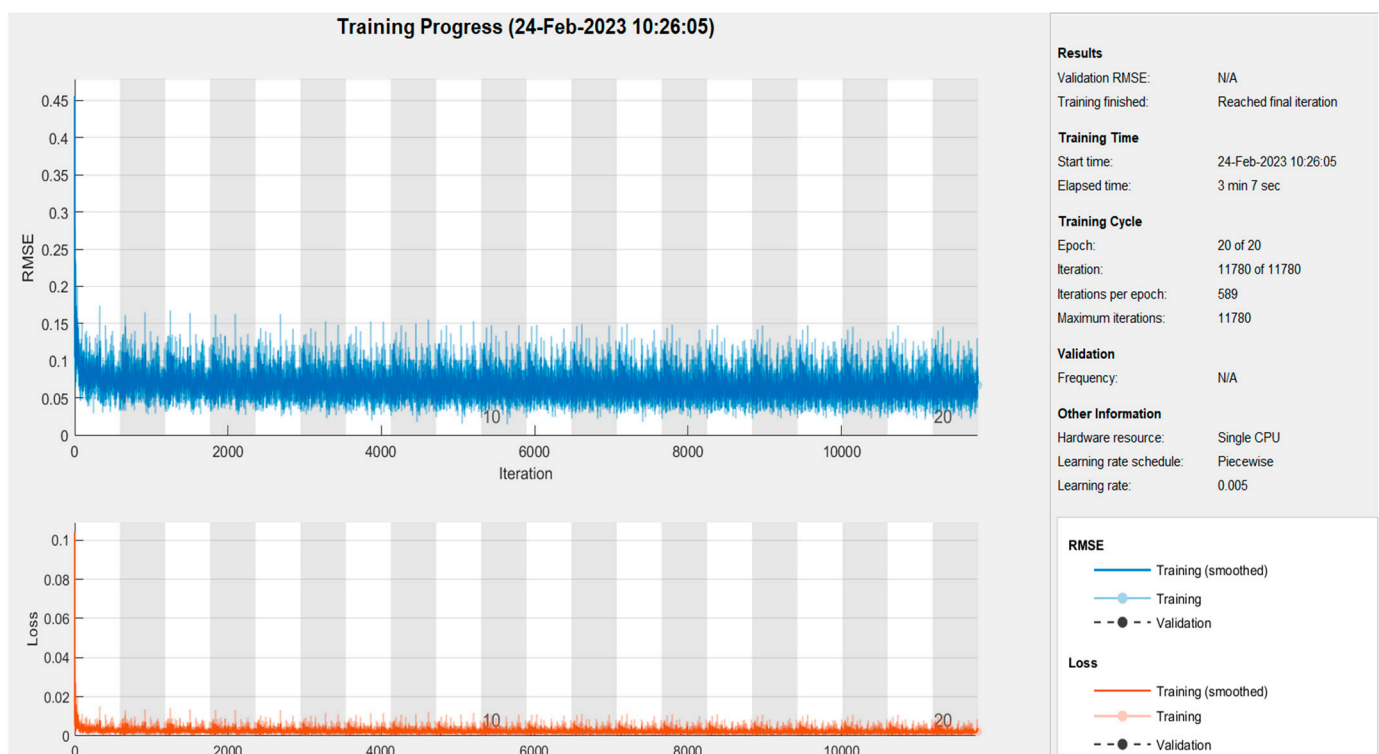


Figure 13. Training process of the bidirectional long short-term memory network model.

3.1. Training Results

The training procedure is a key step in the process of developing a highly effective model by employing certain experimental data. Approximately 80% of the datasets were used for this purpose throughout this stage of the process. Figure 14 illustrates the performance of both the created LSTM and BiLSTM models, and Table 3 presents the values of the assessment metrics. The predicted values of the CO₂% (Y-axis) and the experimental

values (X-axis) are completely consistent across all datasets, as shown in Figure 14 and Table 4, respectively. The constructed LSTM and BiLSTM models are ready to be tested if the BiLSTM has high R% (97.07%) and R2 (93.78) values, as well as extremely low MSE and RMSE values. These values indicate that the system can achieve specified goals.

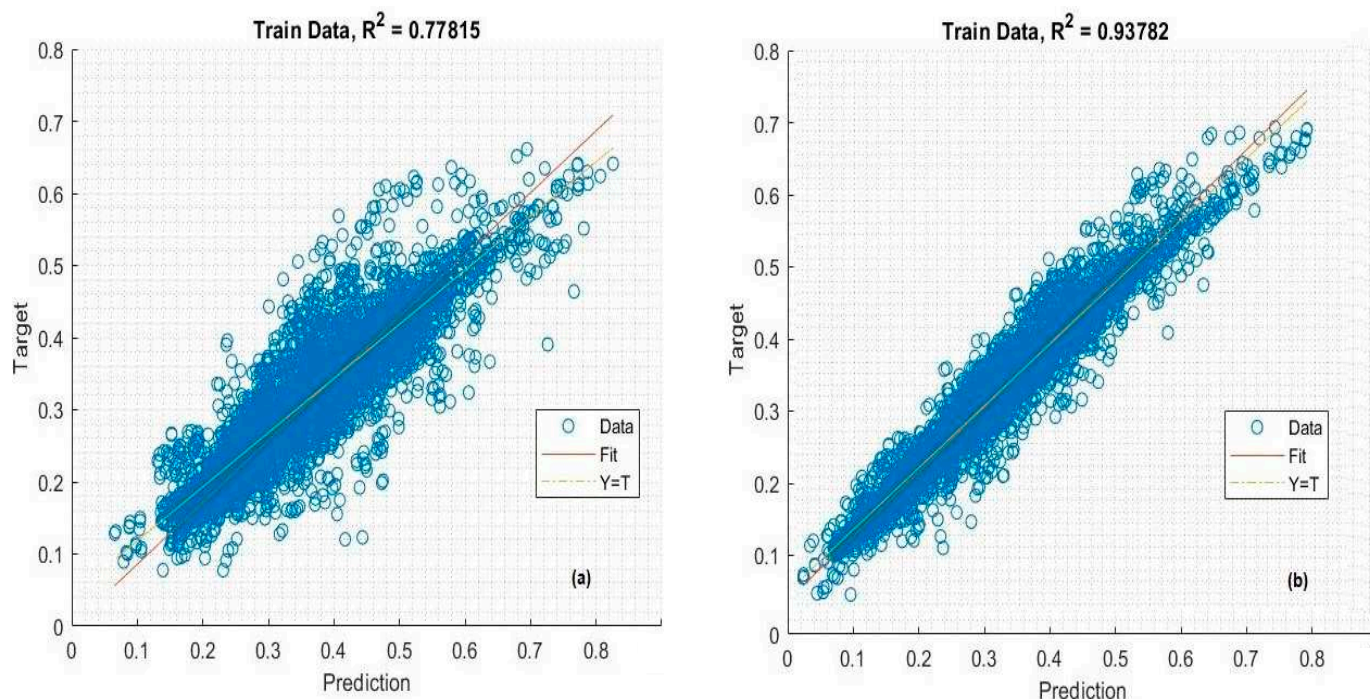


Figure 14. Regression plot of training carbon dioxide (a) LSTM (b) BiLSTM models.

Table 4. Results of LSTM and BiLSTM models in predicting carbon dioxide emissions at the training phase.

Models	#MSE	#RMSE	#R (%)	R ² (%)
LSTM model	0.004980	0.07057	90.47	77.81
BiLSTM model	0.001177	0.0343	97.07	93.78

3.2. Testing Results

Unseen data from 20% of the datasets were used during the testing phase to validate the LSTM and BiLSTM models. Figure 12 and Table 5 shows the testing results for both the LSTM and BiLSTM models, respectively. Figure 15 indicates a strong correlation between the predicted values and expected values (experimental). In addition, the MSE and RMSE values were extraordinarily low at 0.0012 and 0.0035, respectively, whereas the R% and R² values were extremely high at 97.07% and 0.93.78, respectively. The accuracy of the BiLSTM model to predict CO₂ emissions can aid the government in making decisions to prevent the increase of pollution. Our results show that the BiLSTM model has the highest degree of precision in forecasts.

Table 5. Results of LSTM and BiLSTM models in predicting carbon dioxide emissions at the testing phase.

Models	#MSE	#RMSE	#R (%)	#R ² (%)
LSTM model	0.005075	0.07125	90.14	75.73
BiLSTM model	0.0012678	0.03560	96.95	93.55

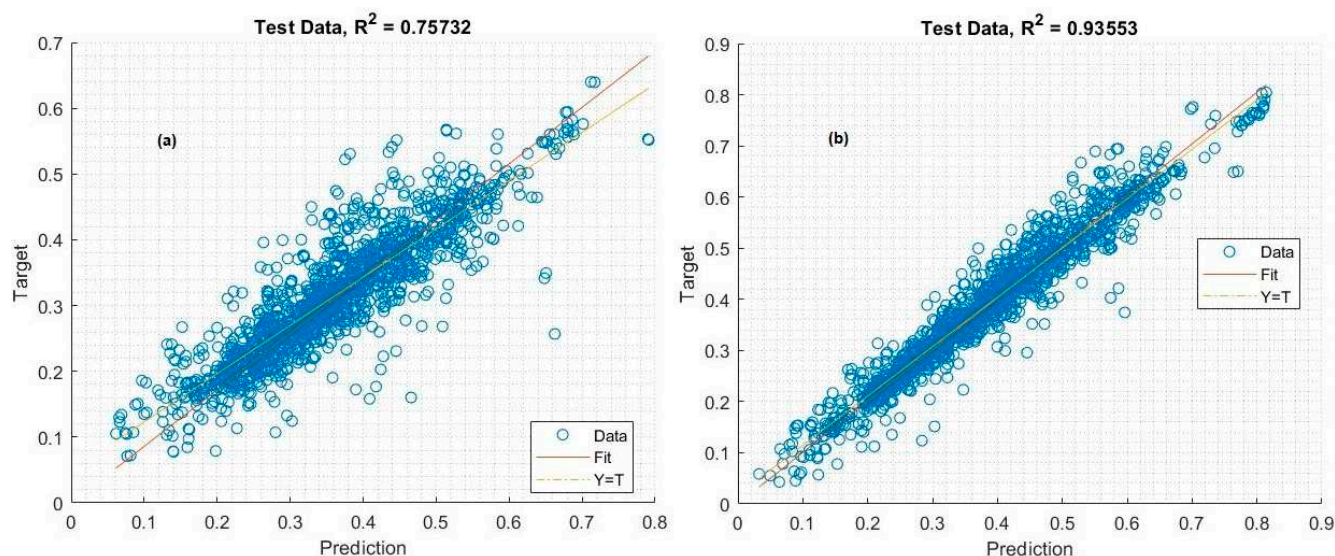


Figure 15. Regression plot of testing carbon dioxide (a) LSTM (b) BiLSTM.

4. Discussion

Emissions from motor vehicles continue to have a substantial influence on both the climate and the quality of the air we breathe. In many regions worldwide, the regulations governing the testing of light- and heavy-duty vehicles are currently being reviewed, and tight emission limits for pollutants, as well as climate policies and targets, are currently being developed.

When it comes to selecting the most effective strategies to cut CO₂ emissions, having an accurate prediction of those emissions is important. Because they take into account the CO₂ produced by a variety of vehicle types, the models that were developed as part of this research are significantly more accurate at predicting CO₂ emissions than any of the existing models, despite the fact that some of those models already exist and have already been developed.

Given the toll it takes on human life and the built environment, pollution is recognized as a major problem worldwide. For this reason, accurate modeling strategies are required to forecast CO₂ emissions in highly populated places, such as megacities. Studies on CO₂ emissions have recently shown substantial promise from the use of deep learning algorithms, such as LSTM and BiLSTM. In this article, we provided a novel technique for estimating CO₂ emissions from vehicles based on deep learning. Our approach has the potential to attain both high usage accuracy and practicality. Data from different types of vehicles and operating parameters were obtained from the onboard diagnostic interface and were used to make predictions regarding exhaust emissions. These parameters included the following: vehicle class, engine_size (L), number of cylinders, transmission type, fuel type, fuel_consumption_city (L/100 km), fuel_consumption_hwy (L/100 km), fuel_consumption_city_comb (L/100 km), fuel_consumption_city_comb (mpg), and CO₂ emissions (g/km). The relative importance of each parameter to emission prediction was analyzed in detail by comparing and contrasting the R². The model's accuracy in estimating CO₂ emissions is proportional to the number of elements utilized as inputs. The more components, the better. Despite this, the amount of accuracy increase that may be accomplished is determined by the parameters for the input.

Figures 16 and 17 show the histogram error of training and testing for predicting CO₂ emissions. The mean errors for the LSTM and BiLSTM models in the training phase were 0.0000568 and 0.00399, respectively. The mean histogram errors of the LSTM and BiLSTM models at the testing phase were 0.0457 and 0.0457, respectively. Finally, this finding indicates that the BiLSTM model achieved fewer prediction errors. The results of

a comparison between the proposed system BiLSTM and several prediction models are shown in Table 6.

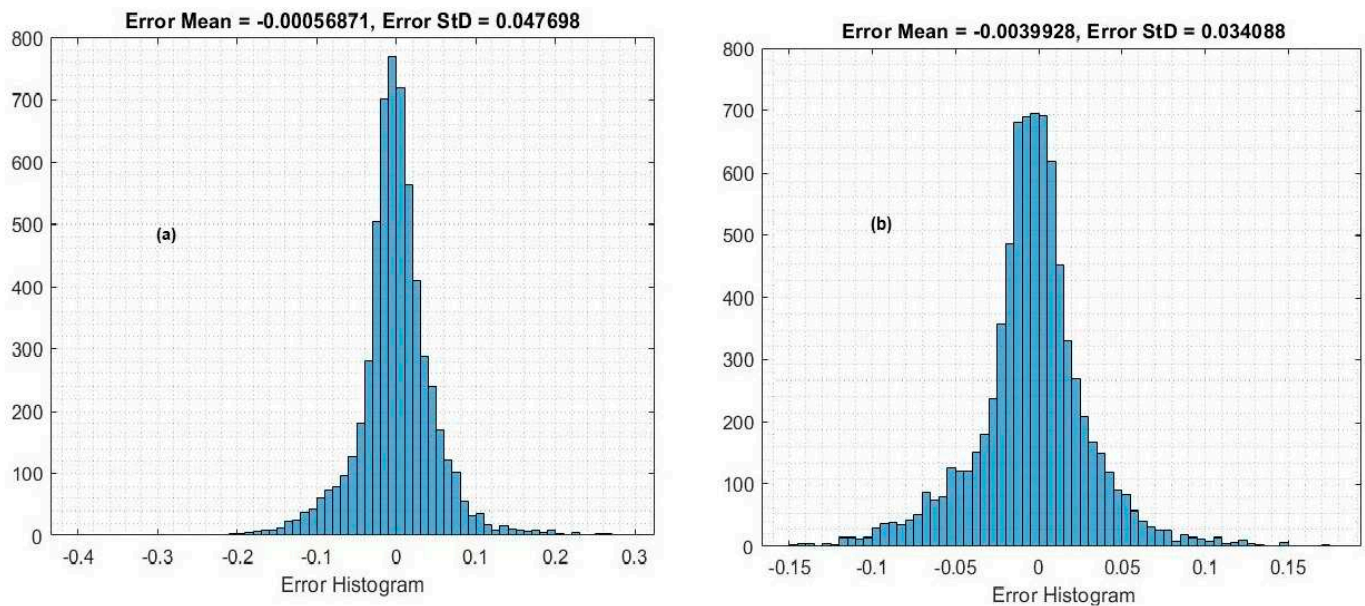


Figure 16. Histogram error at the training step (a) LSTM (b) Bi-LSTM.

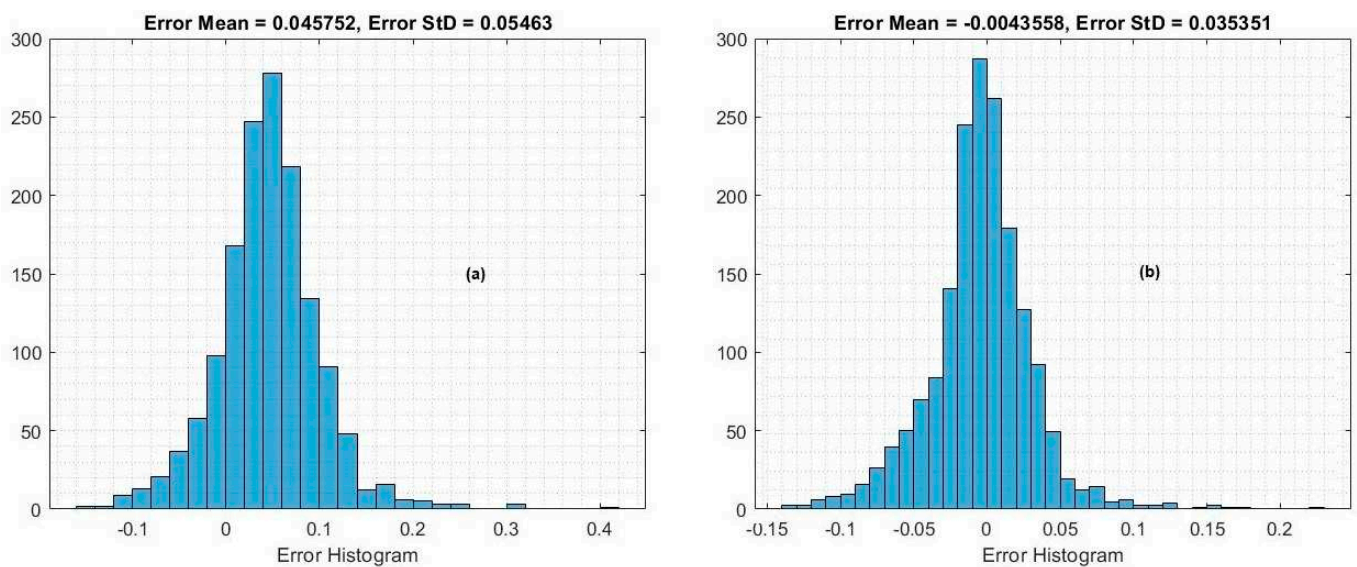


Figure 17. Histogram error at the testing step (a) LSTM (b) Bi-LSTM.

Table 6. Comparison between the results of the estimation CO_2 emission and our model.

Ref	Model	RMSE
Ref. [57]	Artificial Neural Network	1.286
	SVR	2.752
Ref. [58]	long short-term memory network	0.1648
Ref. [59]	Artificial Neural Network	2.95
Ref. [60]	Support Vector Regressor (SVR)	0.71
Ref. [61]	Random forest tree	MAE = 0.58
Proposed	BiLSTM	0.03560

5. Conclusions

Increasing levels of CO₂ emissions have been connected to various adverse consequences on human health, both directly and indirectly. When breathed in high quantities, it can potentially induce severe diseases, including dyspnea, blindness, vertigo, and even delirium. A high level of CO₂ emission contributes to the development of global problems, such as climate change, acid rain, and global warming. Therefore, modeling and predicting CO₂ emission is very important for forecasting future values of CO₂ emission. The following inferences may be made about the modeling and forecast of CO₂ emissions from motor vehicles:

- The development of intelligence system prediction models based on artificial intelligence algorithms for vehicle CO₂ emissions is essential for public health and for the government's capacity to act, CO₂ emissions over the next several decades should be accurately predicted.
- According to the available research, the consequences of CO₂ emissions are observed in numerous countries. One of the things that we can do to increase humans' survivability is to reduce the amount of damaging CO₂ emissions. To investigate the pattern of CO₂ emissions, we employed a time series model based on deep learning.
- After the model's performance regarding the essential performance indicators is analyzed, a suitable model was selected for future forecasting. According to the findings of recent studies, the LSTM and BiLSTM models provide the greatest performance when compared with other models that may be used for precise CO₂ emission prediction.
- According to the findings of recent studies, the LSTM and BiLSTM models provide the greatest performance when compared with other models that may be used for precise CO₂ emission prediction. To assess the performance of the models, we used 7 years' worth of data regarding CO₂ emissions. The most successful model was the BiLSTM, which had an MSE of 0.00126, an RMSE of 0.03560, and an R² of 93.35%. These figures were derived from the values observed for the performance metrics. Using this method, we can make predictions regarding CO₂ emissions in a single dimension.
- Numerous facets have not yet been engrained, including future governmental actions, the transition to renewable energy sources, and economic progress, which are some examples of external factors that may be examined in the future.

Author Contributions: T.H.H.A. and A.H.A.-N. resources, T.H.H.A. data curation, A.H.A.-N. and A.H.A.-N.; writing—original draft preparation, T.H.H.A. and A.H.A.-N. writing—review and editing, A.H.A.-N.; visualization, T.H.H.A. and A.H.A.-N. supervision, T.H.H.A.; project administration, T.H.H.A. and A.H.A.-N.; funding acquisition, T.H.H.A. and A.H.A.-N. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number INST069.

Data Availability Statement: <https://www.kaggle.com/datasets/debajyotipodder/co2-emission-by-vehicles> (last accessed December 2022).

Acknowledgments: This research and the APC were funded by Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number INST069.

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

References

1. Klemm, C.; Wiese, F. Indicators for the optimization of sustainable urban energy systems based on energy system modeling. *Energy Sustain. Soc.* **2022**, *12*, 3. [CrossRef]
2. Sajede, A.; Mohsen, S.; Fahime, M.; Borna, A. Factors affecting the emission of pollutants in different types of transportation. *Energy Rep.* **2022**, *8*, 2508–2529.
3. Menendez, M.; Ambühl, L. Implementing Design and Operational Measures for Sustainable Mobility: Lessons from Zurich. *Sustainability* **2022**, *14*, 625. [CrossRef]
4. Guo, Y.; Ma, Z.; Ren, B.; Zhao, B.; Liu, P.; Zhang, J. Effects of Humic Acid Added to Controlled-Release Fertilizer on Summer Maize Yield, Nitrogen Use Efficiency and Greenhouse Gas Emission. *Agriculture* **2022**, *12*, 448. [CrossRef]

5. Fang, D.; Mueller, C. Mortise-and-tenon joinery for modern timber construction: Quantifying the embodied carbon of an alternative structural connection. *Arch. Struct. Constr.* **2021**, *3*, 11–24. [\[CrossRef\]](#)
6. Górka, M.; Bezyk, Y.; Sówka, I. Assessment of GHG Interactions in the Vicinity of the Municipal Waste Landfill Site—Case Study. *Energies* **2021**, *14*, 8259. [\[CrossRef\]](#)
7. Stubenrauch, J.; Garske, B.; Ekardt, F.; Hagemann, K. European Forest Governance: Status Quo and Optimizing Options with Regard to the Paris Climate Target. *Sustainability* **2022**, *14*, 4365. [\[CrossRef\]](#)
8. IEA (International Energy Agency). Data and Statistics: “CO₂ Emissions by Sector, World 1990–2019”. Available online: <https://www.iea.org/data-and-statistics/data-browser?country=WORLD&fuel=CO2%20emissions&indicator=CO2BySector> (accessed on 10 February 2023).
9. ProAire. Programa Para Mejorar la Calidad del Aire en Mexicali 2011–2020. Available online: https://www.gob.mx/cms/uploads/attachment/file/69289/12_ProAire_Mexicali.pdf (accessed on 10 February 2023).
10. Zhang, H.; Mu, J.E.; McCarl, B.A.; Yu, J. The impact of climate change on global energy use. *Mitig. Adapt. Strat. Glob. Chang.* **2022**, *27*, 9. [\[CrossRef\]](#)
11. Garrido, P. CO₂ Emissions arising from the displacement of the population in private transport mode in Gran Santiago. *Rev. Geogr. Espac.* **2013**, *3*, 69–86.
12. Obaid, M.; Torok, A. Macroscopic Traffic Simulation of Autonomous Vehicle Effects. *Vehicles* **2021**, *3*, 187–196. [\[CrossRef\]](#)
13. Bastien, L.A.; McDonald, B.C.; Brown, N.J.; Harley, R.A. High-resolution mapping of sources contributing to urban air pollution using adjoint sensitivity analysis: Benzene and diesel black carbon. *Environ. Sci. Technol.* **2015**, *49*, 7276–7284. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Fameli, K.M.; Assimakopoulos, V.D. Development of a road transport emission inventory for Greece and the Greater Athens Area: Effects of important parameters. *Sci. Total Environ.* **2015**, *505*, 770–786. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Borge, R.; Narros, A.; Artinano, B.; Yagüe, C.; Gomez-Moreno, F.J.; de la Paz, D.; Quaassdorff, C. Assessment of microscale spatio-temporal variation of air pollution at an urban hotspot in Madrid (Spain) through an extensive field campaign. *Atmos. Environ.* **2016**, *140*, 432–445. [\[CrossRef\]](#)
16. Oftedal, B.; Krog, N.H.; Pyko, A.; Eriksson, C.; Graff-Iversen, S.; Haugen, M.; Aasvang, G.M. Road traffic noise and markers of obesity—a population-based study. *Environ. Res.* **2015**, *138*, 144–153. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Ancona, C.; Badaloni, C.; Mattei, F.; Cesaroni, G.; Stafoggia, M.; Forastiere, F. Health Impact Assessment of Air Pollution, Noise, and Lack of Green in Rome. *J. Transp. Health* **2017**, *5*, S42–S43. [\[CrossRef\]](#)
18. Garshick, E.; Laden, F.; Hart, J.E.; Caron, A. Residence near a major road and respiratory symptoms in US veterans. *Epidemiology* **2003**, *14*, 728. [\[CrossRef\]](#)
19. Delfino, R.J.; Tjoa, T.; Gillen, D.L.; Staimer, N.; Polidori, A.; Arhami, M.; Longhurst, J. Traffic-related air pollution and blood pressure in elderly subjects with coronary artery disease. *Epidemiology* **2010**, *21*, 396–404. [\[CrossRef\]](#)
20. Crouse, D.L.; Goldberg, M.S.; Ross, N.A.; Chen, H.; Labrèche, F. Postmenopausal breast cancer is associated with exposure to traffic-related air pollution in Montreal, Canada: A case-control study. *Environ. Health Perspect.* **2010**, *118*, 1578. [\[CrossRef\]](#)
21. hong, S.; Zhang, K.; Bagheri, M.; Burken, J.G.; Gu, A.; Li, B.; Ma, X.; Marrone, B.L.; Ren, Z.J.; Schrier, J.; et al. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environ. Sci. Technol.* **2021**, *55*, 12741–12754.
22. Junepyo, C.; Park, J.; Lee, H.; Chon, M.S. A Study of Prediction Based on Regression Analysis for Real-World CO₂ Emissions with Light-Duty Diesel Vehicles. *Int. J. Automot. Technol.* **2021**, *22*, 569–577.
23. Aldhyani, T.H.H.; Alkahtani, H. A Bidirectional Long Short-Term Memory Model Algorithm for Predicting COVID-19 in Gulf Countries. *Life* **2021**, *11*, 1118. [\[CrossRef\]](#)
24. Hashemi, N.; Clark, N.N. Artificial Neural Network as a Predictive Tool for Emissions from Heavy-Duty Diesel Vehicles in Southern California. *Int. J. Engine Res.* **2007**, *8*, 321–336. [\[CrossRef\]](#)
25. Jigu, S.; Yun, B.; Park, J.; Park, J.; Shin, M.; Park, S. Prediction of Instantaneous Real-World Emissions from Diesel Light-Duty Vehicles Based on an Integrated Artificial Neural Network and Vehicle Dynamics Model. *Sci. Total Environ.* **2021**, *786*, 147359.
26. Pai, P.S.; Rao, B.S. Artificial Neural Network Based Prediction of Performance and Emission Characteristics of a Variable Compression Ratio CI Engine Using WCO as a Biodiesel at Different Injection Timings. *Appl. Energy* **2011**, *88*, 2344–2354.
27. Cai, M.; Yin, Y.; Xie, M. Prediction of Hourly Air Pollutant Concentrations near Urban Arterials Using Artificial Neural Network Approach. *Transp. Res. Part D Transp. Environ.* **2009**, *14*, 32–41. [\[CrossRef\]](#)
28. Jaikumar, R.; Nagendra, S.S.; Sivanandan, R. Modeling of Real Time Exhaust Emissions of Passenger Cars Under heterogeneous Traffic Conditions. *Atmos. Pollut. Res.* **2016**, *8*, 80–88. [\[CrossRef\]](#)
29. Singh, D.; Kumar, A.; Kumar, K.; Singh, B.; Mina, U.; Singh, B.B.; Jain, V.K. Statistical modeling of O₃, NO_x, CO, PM_{2.5}, VOCs and noise levels in commercial complex and associated health risk assessment in an academic institution. *Sci. Total Environ.* **2016**, *572*, 586–594. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Behera, S.N.; Sharma, M.; Mishra, P.K.; Nayak, P.; Damez-Fontaine, B.; Tahon, R. Passive measurement of NO₂ and application of GIS to generate spatially-distributed air monitoring network in urban environment. *Urban Clim.* **2015**, *14*, 396–413. [\[CrossRef\]](#)
31. Johnson, M.; Isakov, V.; Touma, J.S.; Mukerjee, S.; Özkaynak, H. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmos. Environ.* **2010**, *44*, 3660–3668. [\[CrossRef\]](#)
32. Kanaroglou, P.S.; Adams, M.D.; De Luca, P.F.; Corr, D.; Soheli, N. Estimation of sulfur dioxide air pollution concentrations with a spatial autoregressive model. *Atmos. Environ.* **2013**, *79*, 421–427. [\[CrossRef\]](#)

33. Vandaele, N.; Van Woensel, T.; Verbruggen, A. A queueing based traffic flow model. *Transp. Res. Part D Transp. Environ.* **2000**, *5*, 121–135. [\[CrossRef\]](#)
34. Tomić, J.; Bogojević, N.; Pljakić, M.; Šumarac-Pavlović, D. Assessment of traffic noise levels in urban areas using different soft computing techniques. *J. Acoust. Soc. Am.* **2016**, *140*, EL340–EL345. [\[CrossRef\]](#)
35. Hamad, K.; Khalil, M.A.; Shanableh, A. Modeling roadway traffic noise in a hot climate using artificial neural networks. *Transp. Res. Part D. Transp. Environ.* **2017**, *53*, 161–177. [\[CrossRef\]](#)
36. Di Mascio, P.; Di Vito, M.; Loprencipe, G.; Ragnoli, A. Procedure to determine the geometry of road alignment using GPS data. *Procedia-Soc. Behav. Sci.* **2012**, *53*, 1202–1215. [\[CrossRef\]](#)
37. Hamrani, A.; Akbarzadeh, A.; Madramootoo, C.A. Machine Learning for Predicting Greenhouse Gas Emissions from Agricultural Soils. *Sci. Total Environ.* **2020**, *741*, 140338. [\[CrossRef\]](#)
38. Tao, J.; Zhang, P.; Chen, B. A Microscopic Model of Vehicle CO₂ Emissions Based on Deep Learning—A Spatiotemporal Analysis of Taxicabs in Wuhan, China. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 18446–18455.
39. Wang, A.; Xu, J.; Zhang, M.; Zhai, Z.J.; Song, G.; Hatzopoulou, M. Emissions and fuel consumption of a hybrid electric vehicle in real-world metropolitan traffic conditions. *Appl. Energy* **2022**, *306*, 118077. [\[CrossRef\]](#)
40. Singh, M.; Dubey, R. Deep Learning Model Based CO₂ Emissions Prediction using Vehicle Telematics Sensors Data. *IEEE Trans. Intell. Veh.* **2021**, *8*, 768–777. [\[CrossRef\]](#)
41. Mądziel, M.; Jaworski, A.; Kuszewski, H.; Woś, P.; Campisi, T.; Lew, K. The Development of CO₂ Instantaneous Emission Model of Full Hybrid Vehicle with the Use of Machine Learning Techniques. *Energies* **2021**, *15*, 142. [\[CrossRef\]](#)
42. Xie, H.; Zhang, Y.; He, Y.; You, K.; Fan, B.; Yu, D.; Lei, B.; Zhang, W. Parallel attention-based LSTM for building a prediction model of vehicle emissions using PEMS and OBD. *Measurement* **2021**, *185*, 110074. [\[CrossRef\]](#)
43. Shahnavaaz, F.; Akhavan, R. Automated Estimation of Construction Equipment Emission Using Inertial Sensors and Machine Learning Models. *Sustainability* **2022**, *14*, 2750. [\[CrossRef\]](#)
44. Xayasouk, T.; Lee, H.; Lee, G. Air Pollution Prediction Using Long Short-Term Memory (LSTM) and Deep Autoencoder (DAE) Models. *Sustainability* **2020**, *12*, 2570. [\[CrossRef\]](#)
45. Wang, X.; Yan, C.; Liu, W.; Liu, X. Research on Carbon Emissions Prediction Model of Thermal Power Plant Based on SSA-LSTM Algorithm with Boiler Feed Water Influencing Factors. *Sustainability* **2022**, *14*, 15988. [\[CrossRef\]](#)
46. Gu, B.; Zhang, T.; Meng, H.; Zhang, J. Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation. *Renew. Energy* **2021**, *164*, 687–708. [\[CrossRef\]](#)
47. Sakshi, K.V.A. An ARIMA-LSTM hybrid model for stock market prediction using live data. *J. Eng. Sci. Technol. Rev.* **2020**, *13*, 117–123. [\[CrossRef\]](#)
48. Dave, E.; Leonardo, A.; Jeanive, M.; Hanafiah, N. Forecasting Indonesia exports using a hybrid model ARIMA-LSTM. *Procedia Comput. Sci.* **2021**, *179*, 480–487. [\[CrossRef\]](#)
49. Huang, S.; Shen, J.; Lv, Q.; Zhou, Q.; Yong, B. A Novel NODE Approach Combined with LSTM for Short-Term Electricity Load Forecasting. *Future Internet* **2023**, *15*, 22. [\[CrossRef\]](#)
50. Al-Nefaie, A.H.; Aldhyani, T.H.H. Predicting Close Price in Emerging Saudi Stock Exchange: Time Series Models. *Electronics* **2022**, *11*, 3443. [\[CrossRef\]](#)
51. Lingras, P.; West, C. Interval set clustering of web users with rough k-means. *J. Intell. Inf. Syst.* **2004**, *23*, 5–16. [\[CrossRef\]](#)
52. Aldhyani, T.H.; Joshi, M.R. Handling ambiguous packets in intrusion detection. In Proceedings of the 3rd International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, India, 26–28 March 2015.
53. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *Appl. Stat.* **1979**, *28*, 100–108. [\[CrossRef\]](#)
54. Wada, K. Outliers in official statistics. *Jpn. J. Stat. Data Sci.* **2020**, *3*, 669–691. [\[CrossRef\]](#)
55. Al-Nefaie, A.H.; Aldhyani, T.H.H. Bitcoin Price Forecasting and Trading: Data Analytics Approaches. *Electronics* **2022**, *11*, 4088. [\[CrossRef\]](#)
56. Al-Adhaileh, M.H.; Aldhyani, T.H. Artificial intelligence framework for modeling and predicting crop yield to enhance food security in Saudi Arabia. *Peer J. Comput. Sci.* **2022**, *8*, e1104. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Azeez, O.S.; Pradhan, B.; Shafri, H.Z.M.; Shukla, N.; Lee, C.-W.; Rizzei, H.M. Modeling of CO Emissions from Traffic Vehicles Using Artificial Neural Networks. *Appl. Sci.* **2019**, *9*, 313. [\[CrossRef\]](#)
58. Zhang, R.; Wang, Y.; Pang, Y.; Zhang, B.; Wei, Y.; Wang, M.; Zhu, R. A Deep Learning Micro-Scale Model to Estimate the CO₂ Emissions from Light-Duty Diesel Trucks Based on Real-World Driving. *Atmosphere* **2022**, *13*, 1466. [\[CrossRef\]](#)
59. Debone, D.; Martins, T.D.; Miraglia, S.G.E.K. Modeling Carbon Release of Brazilian Highest Economic Pole and Major Urban Emitter: Comparing Classical Methods and Artificial Neural Networks. *Climate* **2022**, *10*, 9. [\[CrossRef\]](#)
60. Chadha, A.S.; Shinde, Y.; Sharma, N.; De, P.K. Predicting CO₂ Emissions by Vehicles Using Machine Learning. In *Data Management, Analytics and Innovation*; Goswami, S., Barara, I.S., Goje, A., Mohan, C., Bruckstein, A.M., Eds.; ICDMAI 2022. Lecture Notes on Data Engineering and Communications Technologies; Springer: Singapore, 2023; Volume 137. [\[CrossRef\]](#)
61. Shah, S.; Thakar, S.; Jain, K.; Shah, B.; Dhagel, S. A Comparative Study of Machine Learning and Deep Learning Techniques for Prediction of Co₂ Emission in Cars. *arXiv* **2022**, arXiv:2211.08268.

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.