Climate Change Reports

Journal Homepage: sciencesforce.com/ccr

Climate Cha. Rep. Vol. 1 (2024) 19–29

Paper Type: Original Article

Application of Deep Learning Initiatives for CO2 Emissions Forecasting

Nihal N. Mostafa [1], Ahmed Tolba [2], and Mohamed Abouhawwash [3,*]

[1] Department of Computer Science, Misr higher institute for computer and commerce, Egypt; nihal.nabil@fci.zu.edu.eg.
[2] Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt; a.tolba24@fci.zu.edu.eg.
[3] Department of Computational Mathematics, Science, and Engineering (CMSE), Michigan State University, East Lansing, United States; abouhaww@msu.edu.

Received: 06 Sep 2023 **Revised:** 02 Jan 2024 **Accepted:** 27 Jan 2024 **Published:** 29 Jan 2024

Abstract

This work models and forecasts vehicle CO2 emissions, a major source of atmospheric changes and climate disruptions, using cutting-edge artificial intelligence. The CO2 emission by vehicle dataset from Kaggle, which includes several features such vehicle class, engine size, cylinder transmission, fuel type, fuel consumption, city, highway, comb, and CO2 emissions, was used to build the model. To predict CO2 emissions, a hybrid model (CNN-LSTM-MLP) was developed based on long short-term memory network (LSTM), convolution neural network (CNN), and multi-layer perceptron (MLP). The proposed model shows superior results compared with CNN, MLP, LSTM, Light Gradient Boosting Machine (LGBM) Regressor, support vector machine (SVM), Linear Regression, and Random Forest.

Keywords: Deep Learning; CO2 Emissions; Long Short-Term Memory Network; Convolution Neural Network.

1 | Introduction

Climate change-related disasters have caused 77% of global economic losses between 1998 and 2017. The world is accelerating the transition to low-carbon energy systems, but progress lags behind the Paris climate targets. The industrial sector accounts for 37% of global energy consumption and 18.4% of global carbon emissions [1]. Despite automation, manual decision-making leads to wasted energy and carbon emissions. Accurate forecasting of energy consumption and emissions is crucial for optimizing production scheduling, optimizing energy utilization, and reducing energy and material waste. Proposing energy consumption and carbon emissions forecasting models can help improve production planning and energy management systems [2].

Both positive and negative effects of technology can be seen in CO2 emissions. Intelligent technology and techniques have advanced quickly in recent years, and scholars and practitioners have looked more closely at how they might be applied in the construction industry to reduce emissions and promote sustainability. The construction sector might potentially reduce CO2 emissions by implementing automated procedures, new

Corresponding Author: abouhaww@msu.edu
https://doi.org/10.61356/j.ccr.2024.1209
Licensee **Climate Change Reports**. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0).

materials, and technologies that replace existing ones in both hardware and soft computing. The world is moving toward novel procedures including machine learning (ML), Deep learning (DL), real-time monitoring, Internet of Things (IoT), and optimization techniques [3].

Moon et al. [4] utilized real-driving emission tests on commercial vehicles to gather CO2 emissions and engine control unit data. A ML model, XGBoost, is trained to process the data. The method, based on predictive CO2 emissions monitoring, shows greater accuracy than a fuel-based method, emphasizing the need for adoption. Unfortunately, ML-based model extract features manually affect the overall accuracy of the model classification task.

For public buildings in big spaces, Wang and Hu [5] analyzed four ML-based generators using a year's worth of recorded data to obtain the best possible thermal comfort management and carbon emission forecast. The hybrid ML approach can minimize carbon emissions more than manual management, according to the optimization results. A mixed prediction model is put forth by Yang and Wang [6] to forecast carbon emissions. The combined model first split the original data into two groups, then trained kernel parameters using the carnivorous plant and chameleon swarm algorithms, and lastly used the model to forecast carbon emissions. Every index of this model was shown to be superior to those of other comparative models.

Al-Nefaie et al. [7] proposed advanced DL to model and predict vehicle CO2 emissions using a Kaggle dataset. Two models, a LSTM, and a bidirectional LSTM (BiLSTM), were developed. The BiLSTM model performed best, achieving high prediction values for MSE and RMSE.

A Bayesian neural network model (IO-BNN) based on the economic input-output technique was proposed by Zhou and Li[8]. Using this model, the authors analyze Guangdong's carbon emissions; the findings indicate that, in the best-case scenario for policy, Guangdong's carbon emissions should peak in 2025.

Previous studies suffer from some limitations, therefore this study proposed a (CNN-LSTM-MLP) Hybrid model. The proposed model aims to overcome the previous studies' limitations. The main contribution of this study is discussed as follows:

- The proposed (CNN-LSTM-MLP) Hybrid model can capture both local and global features, manage sequential or temporal dependencies, and handle diverse forms of input.
- LSTM and CNN in the proposed model can extract spatial and temporal patterns from CO2 emissions data.
- MLP aims to analyze and capture complex relationships and give final output.
- The proposed (CNN-LSTM-MLP) model shows superior results compared with CNN, MLP, LSTM, Light Gradient Boosting Machine (LGBM) Regressor, support vector machine (SVM), Linear Regression, and Random Forest.

The remainder of the paper is divided as follows. Section 2 presents the material and method of this study. Section 3 presents the proposed model. Section 4 presents experimental results. Section 5 illustrates the conclusion and future directions of this proposal.

2 | Materials and Methods

2.1 | Dataset

The dataset CO_2 Emissions involves 11 unique features {Make, 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)}. These criteria are employed to predict the CO2 Emissions(g/km) [9] as shown in Table 1.

Table 1. Dataset features.

Categorical features	Numerical features
'Make' has 42 unique values	Engine Size(L)
'Model ' has 2053 unique values	Cylinders
'Vehicle' has 16 unique values	Fuel Consumption City
'Transmission' has 27 unique values	Fuel Consumption Hwy (L/100 km)
'Fuel Type' has 5 unique values	Fuel Consumption Comb (L/100 km)
	Fuel Consumption Comb (mpg)

2.2 | Dataset Distribution

Used different statistical measures such as mean, standard deviation, and quartiles (minimum, 25th percentile, median, 75th percentile, maximum) to provide more insights into the central tendency, spread, and distribution of the data. In addition, it enables to identify the outliers and anomalies through the range of values and it's useful for data cleaning, quality assessment, and exploratory data analysis to use in predictions.

The Mean and Standard Deviation (std) of these values are calculated by the following equation respectively:

$$\mu = \frac{\sum x_i}{n} \tag{1}$$

Where:

- μ = Mean
- xi = ith observation, $1 \leq i \leq n$
- $\sum$xi = Sum of observations
- n = Number of observations

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{n}} \tag{2}$$

Where:

- σ = Standard Deviation
- xi = Terms Given in the Data
- μ = Mean
- n = Total number of Terms

Box plots are employed to summarize data distributions, detect skewness, identify outliers, and compare distributions. They offer a graphical depiction of 5 quarters. These are summarized in Figure 1.

- Minimum – The minimum value is the smallest value in the dataset not including any outliers.
- First Quartile (Q1) – The first quartile, also known as the lower quartile, represents the number below which 25% of the data lies.
- The median (Q2) represents the value that divides the dataset into two halves. It divides the values into two equal parts, with half below and half above.
- The Third Quartile (Q3) represents the point at which 75% of the data falls below it.

  Maximum — The highest value in the dataset, not including any outliers.

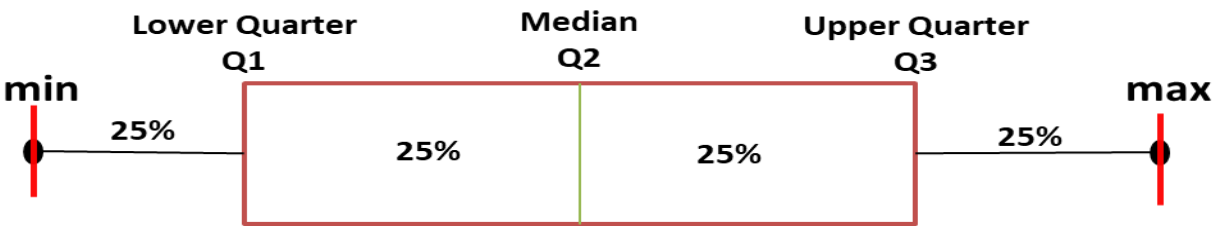


Figure 1. Box plot diagram.

Table 2. Data distributions.

	Count	mean	std	min	25%	50%	75%	Max
Make	7385	19.57021	11.31163	0	9	18	29	41
Model	7385	1023.659	577.0224	0	531	999	1524	2052
Vehicle Class	7385	6.364523	4.822959	0	2	6	11	15
Engine Size(L)	7385	3.160068	1.35417	0.9	2	3	3.7	8.4
Cylinders	7385	5.61503	1.828307	3	4	6	6	16
Transmission	7385	14.02776	7.260507	0	6	15	17	26
Fuel Type	7385	3.262153	0.882482	0	3	3	4	4
Fuel Consumption City (L/100 km)	7385	12.55653	3.500274	4.2	10.1	12.1	14.6	30.6
Fuel Consumption Hwy (L/100 km)	7385	9.041706	2.224456	4	7.5	8.7	10.2	20.6
Fuel Consumption Comb (L/100 km)	7385	10.97507	2.892506	4.1	8.9	10.6	12.6	26.1
Fuel Consumption Comb (mpg)	7385	27.48165	7.231879	11	22	27	32	69
CO2 Emissions(g/km)	7385	250.5847	58.51268	96	208	246	288	522

2.3 | Data Pre-Processing

Initially, verify the presence of null values. It has been determined that there are no missing values in any of the features.

In the Second stage, an examination of the data distribution, the dataset has different distributions As shown in Table 2. Where some variables have a wide range of values with significant variations other variables have smaller ranges and lower variations.

In the Third stage, feature values are scaled using the StandardScaler. In the field of machine learning, the StandardScaler is a preprocessing technique employed to normalize features. This is achieved by eliminating the mean and scaling the features to unit variance. The data is transformed in a manner that assigns a mean of zero and a standard deviation of one to each feature. The transformation performed is expressed mathematically as:

$$Z = \frac{X - \mu}{S} \tag{3}$$

Where Z represents new samples, μ Mean, and S Standard deviation.

Normalization facilitates faster convergence of algorithms and mitigates the dominance of features with larger scales over those with lower scales. The StandardScaler function is utilized to approximate the data to a standard normal distribution.

Finally splitting the dataset, for the DL experiment the dataset is divided into multiple subsets to facilitate training, testing, and validation by 80%, 10 %, and 10%. In the machine learning experiment, the dataset is divided into two parts training and testing by 80% and 20 % respectively.

2.4 |Deep Learning Model

The Proposed (CNN-LSTM-MLP) hybrid model is based combination of CNN, LSTM, and MLP.

MLP stands for Multilayer Perceptron, a specific type of feedforward neural network. The architecture has an input layer, one or more hidden layers, and an output layer [36]. Every layer consists of several interconnected artificial neurons or nodes. MLPs employ non-linear activation functions to incorporate non-linearity into the model, allowing for the acquisition of intricate patterns within the data. The training process involves utilizing backpropagation, a technique in which the error is propagated in a reverse manner across the network to adjust the weights and biases. MLPs are extensively employed for tasks such as classification, regression, and pattern recognition [10].

Long Short-Term Memory (LSTM) is a specific form of recurrent neural network (RNN) that has exceptional proficiency in handling sequential input. The vanishing gradient problem is resolved by integrating memory cells with gating mechanisms. Long Short-Term Memory (LSTM) models can retain information throughout extensive sequences, which makes them well-suited for tasks involving natural language processing and time series analysis. They possess an inherent memory state that grants them the ability to deliberately discard or retain information, thus facilitating the acquisition of long-term connections. LSTMs consist of input, forget, and output gates that regulate the information flow inside the network in addition the activation function is tanh [11].

A convolutional neural network (CNN) is a specialized deep learning model designed specifically for analyzing different datasets, such as images and time series data. Convolutional layers are employed to automatically capture and extract features from the input data. The convolutional layers utilize filters to process the input, enabling the network to detect patterns and spatial relationships. The convolutional layers produce an output which is then sent to fully connected layers to carry out classification or regression tasks. It has shown considerable potential in managing time series data, such as forecasting wind speed, solar irradiance, and stock prices. In the context of 1D convolution, the kernel serves as a filter that extracts features and uses maxpooling to reduce the size of the input representation by selecting the highest value within a specified pool_size window. The activation function used for one-dimensional convolutional feature extraction employs the Rectified Linear Unit (ReLU) activation function, whereas the activation function for the output layer is linear for regression [12].

3 |Proposed Model

The proposed hybrid model is created by integrating various neural network architectures to capitalize on their strengths and tackle unique issues. Hybrid models, which incorporate various models like CNNs, and LSTMs, can capture both local and global features, manage sequential or temporal dependencies, and handle diverse forms of input. Hybrid models utilize the combined strengths of different architectures to improve efficiency and flexibility. This allows for more precise and thorough modeling of intricate data. The Proposed Model Utilizing CNN and LSTM, focuses on extracting spatial and temporal patterns from CO2 emissions data and Utilizing a MLP to analyze and capture complex relationships and give final output. Figure 2 shows the CNN-LSTM-MLP architecture.

The Implementation CNN-LSTM-MLP Model has specific layers and characteristics.

1. Convolutional layers:

- The first Conv1D layer consists of 32 filters with a kernel size of 3, 'relu' activation function, 'same' padding, and the input shape is defined by `input_shape`.

- The MaxPooling1D layer has a pool size of 2.

- The second Conv1D layer consists of 64 filters with a kernel size of 3 and uses 'relu' activation.

The MaxPooling1D layer has a pool size of 2.

2. LSTM layers:

- First LSTM layer with 100 units and returning sequences.

- The second LSTM layer consists of 100 units.

3. Flatten layer:

- Compresses the output prior to forwarding it to the dense layers.

4. MLP layers:

- First dense layer consists of 20 units with 'relu' activation function.

- finally, the dense layer consists of 1 unit.

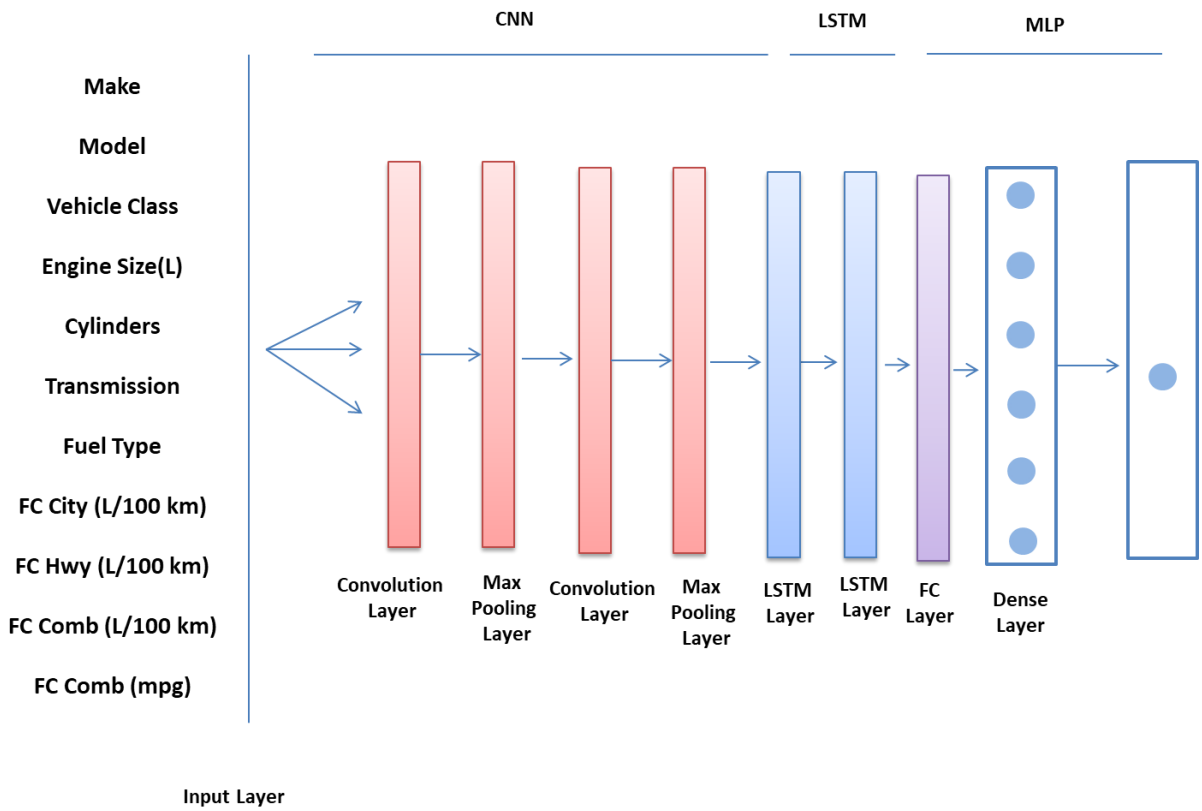


Figure 2. Architecture of the hybrid CNN-LSTM model.

4 | Results and Discussion

This section investigates the performance of the proposed model using a widely used dataset Co2 Emissions. In addition, it is compared to several ML and DL models, such as the LGBM Regressor, Random Forest Regressor, SVM, Linear Regression, MLP, LSTM, CNN, and The Proposed Model. Those DL models are implemented in Python using the Kaggle platform and Keras API. The Adam optimizer was used to train the weights of those models for 150 epochs. ML Models are implemented in Python using the Kaggle platform and scikit learn API and LightGBM library. We have used 8 to evaluate the models' effectiveness. Before using all these metrics, we must know about the residual error, i.e., $(y - \hat{y})$. Here, y and indicate the real

values and $\hat{y}$ indicate predicted values. The performance indicators used to evaluate the performance of those models are:

Mean squared error:

For calculating the MSE, take the real value, subtract the predicted value, and square that difference. Repeat that for all samples. Then, sum all of those squared values and divide by the number of samples.

$$\mathrm{MSE} = \frac{\Sigma(y_i - p_i)^2}{n} \tag{4}$$

Root Mean Square Error

The Root Mean Squared Error (RMSE) is the most common performance indicator for a regression model. It measures the average difference between predicted values and real values. The lower the RMSE, the better the model and its predictions.

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}.(x_i - \hat{x}_i)^2}{N}} \tag{5}$$

Where x_i real values and $\hat{x}_i$ predicted values, N number of samples, $\sum_{i=1}^{N}$. Summation from I to N

Mean Absolute Error

The mean absolute error (MAE) is defined as the average variance between the real and predicted values.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_1 - \hat{y}_i|^2 \tag{6}$$

Where $y_i$predicted value, p_i the real value of I, n number of samples.

R2 Score:

The regression coefficient determines the best possible score is 1.0, and less than that is less efficient and gets worse.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\Sigma\mathrm{i}(y_i - \hat{y}_i)^2}{\Sigma\mathrm{i}(y_i - \overline{y})^2} \tag{7}$$

Where $\hat{y}$ real values and ipredicted values.

Median Absolute Error

The measure of variability in a univariate sample of quantitative data is effectively captured by the median absolute error (MedAE). The value of it might range from 0 to infinity. Hence, a decrease in the value corresponds to an increase in the accuracy of the model.

$$\mathrm{MedAE}(\mathrm{y}, \hat{\mathrm{y}}) = \mathrm{median}(|\mathrm{y}_1 - \hat{\mathrm{y}}_1|, \dots\dots\dots |\mathrm{y}_\mathrm{n} - \hat{\mathrm{y}}_\mathrm{n}|) \tag{8}$$

Max Error

The max error metric calculates the maximum residual error.

$$\mathrm{Max\ error} = \mathrm{Max}\ |\ \mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}\ | \tag{9}$$

Explained Variance Score Error

The calculation of the explained variance score error is utilized to quantify the extent to which the predictions of the machine learning models being applied exhibit variability. Variance quantifies the extent to which

observed values deviate from the mean of expected values, specifically, their deviation from the mean of the predicted values.

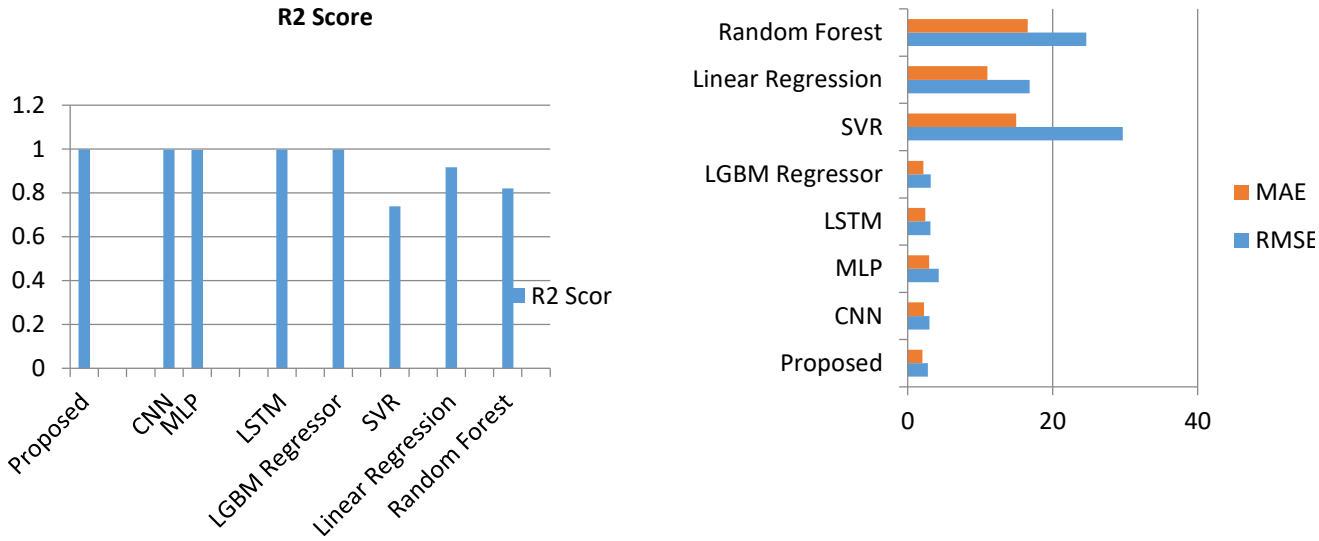
$$\mathrm{EVS}(y, \hat{y}) = 1 - \frac{\mathrm{var}(y - \hat{y})}{\mathrm{var}(y)} \tag{10}$$

Table 3 and Figure 3 show the metrics presented in the table, the proposed model has the highest level of explained variance (0.998) and regression score (0.998), suggesting that it effectively accounts for most of the variance observed in the dependent variable and offers the most optimal fit to the data compared to all other models. In addition, it exhibits the lowest maximum error (19.405), mean squared error (7.747), root mean squared error (2.783), mean absolute error (2.066), and mean average error (1.639), hence showcasing its superior performance in comparison to other models. Based on the measures, it can be inferred that the proposed model exhibits the best performance compared to the other assessed models.

Figure 4 shows graphs depicting the historical loss values and RMSE (Root Mean Square Error) for the proposed model. the data's history across 150 epochs and showcases the loss function values from both the training and validation phases of the model. , the proposed model has outstanding stability and accuracy. The graphs depicting the loss values show steady and slight variations during both the training and validation stages. It starts with loss: 55165.2930 - val_loss: 19159.1953 and ends with loss: 6.6433 - val_loss: 11.4653The stability demonstrates start from epochs 20 and proves that the proposed model effectively learns from the input and generates dependable outcomes. Figure 5 shows how closely the predicted values match the actual values.

Table 3. Performance of proposed CNN-LSTM-MLP model.

Model Name	Explained variance Regression score	Max error	MSE	R2 Score	RMSE	MAE	MedAE
CNN	0.997	16.969	9.076	0.997	3.013	2.284	1.886
MLP	0.995	38.394	18.295	0.995	4.277	2.993	2.356
LSTM	0.997	14.065	9.964	0.997	3.157	2.425	1.936
LGBM Regressor	0.997	32.337	10.270	0.997	3.205	2.196	1.744
SVM	0.739	225.689	882.212	0.739	29.702	14.96	3.895
Linear Regression	0.916	83.512	283.924	0.916	16.850	11.026	6.661
Random Forest	0.820	138.636	607.559	0.820	24.649	16.55	11.871
Proposed	0.998	19.405	7.747	0.998	2.783	2.066	1.639



(a) (b)

Figure 3. CNN-LSTM-MLP results (a) is the performance of the proposed model in terms of R2 score;

(b) is the performance of the proposed model in terms of MAE and RMSE.

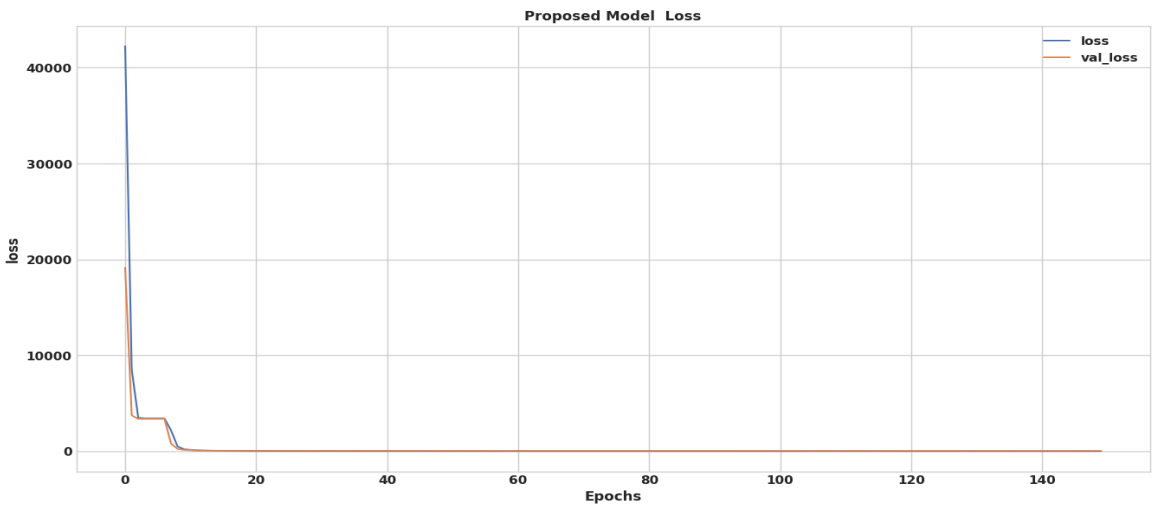


Figure 4. Model history.

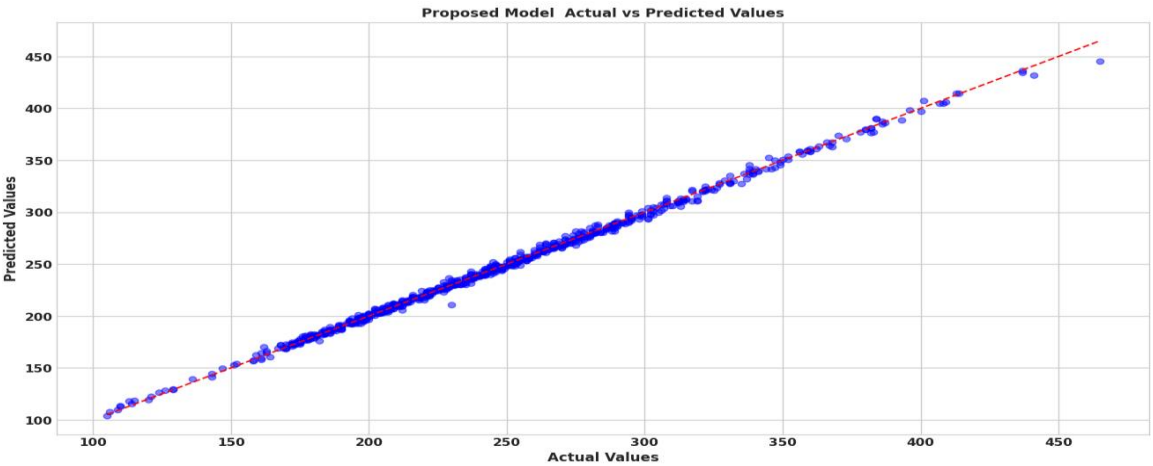


Figure 5. Real value vs predicted value.

5 | Conclusions

Climate change-related disasters cause 77% of global economic losses, impacting industrial sector energy consumption and carbon emissions. Accurate forecasting and intelligent technology can reduce emissions and promote sustainability. In this study we introduce a novel model for vehicle CO2 emissions prediction. The CO2 emission by vehicle dataset from Kaggle, which includes several features such vehicle class, engine size, cylinder transmission, fuel type, fuel consumption, city, highway, comb, and CO2 emissions, was used to build the model. To predict CO2 emissions, a hybrid model (CNN-LSTM-MLP) was developed based on long short-term memory network (LSTM), convolution neural network (CNN), and multi-layer perceptron (MLP). The proposed model shows superior results in terms of explained variance regression score, max error, MSE, R2 Score, RMSE, MAE, and MedAE with 0.998, 19.405, 7.747, 0.998, 2.783, 2.066, 1.639, respectively. Our proposed work was compared with CNN, MLP, LSTM, Light Gradient Boosting Machine (LGBM) Regressor, support vector machine (SVM), Linear Regression, and Random Forest.

Future work should expand the model's application to multiple regions, add socio-economic and industrial variables for CO2 prediction, and develop specialized machine learning or deep learning algorithms for economic growth. Accurate CO2 prediction is crucial for understanding climate change implications, policy decisions, sustainable economic development, and effective adaptation.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Author Contribution

All authors contributed equally to this work.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] Kiss, A.A. and R. Smith, Rethinking energy use in distillation processes for a more sustainable chemical industry. Energy, 2020. 203: p. 117788.

[2] Hu, Y. and Y. Man, Energy consumption and carbon emissions forecasting for industrial processes: Status, challenges and perspectives. Renewable and Sustainable Energy Reviews, 2023. 182: p. 113405.

[3] Sojobi, A. and K. Liew, Multi-objective optimization of high performance bio-inspired prefabricated composites for sustainable and resilient construction. Composite Structures, 2022. 279: p. 114732.

[4] Moon, S., et al., Study on CO2 Emission Assessment of Heavy-Duty and Ultra-Heavy-Duty Vehicles Using Machine Learning. International Journal of Automotive Technology, 2024.
[5] Wang, P., J. Hu, and W. Chen, A hybrid machine learning model to optimize thermal comfort and carbon emissions of large-space public buildings. Journal of Cleaner Production, 2023. 400: p. 136538.
[6] Yang, H., M. Wang, and G. Li, A combined prediction model based on secondary decomposition and intelligence optimization for carbon emission. Applied Mathematical Modelling, 2023. 121: p. 484-505.
[7] Al-Nefaie, A.H. and T.H. Aldhyani, Predicting CO2 emissions from traffic vehicles for sustainable and smart environment using a deep learning model. Sustainability, 2023. 15(9): p. 7615.
[8] Zhou, B., et al., An input-output-based Bayesian neural network method for analyzing carbon reduction potential: A case study of Guangdong province. Journal of Cleaner Production, 2023. 389: p. 135986.
[9] CO2 emission prediction dataset.
[10] MK, A.N. and M.A. V, Role of energy use in the prediction of CO 2 emissions and economic growth in India: evidence from artificial neural networks (ANN). Environmental Science and Pollution Research, 2020. 27: p. 23631-23642.
[11] Cortes, V.M.P., A. Chatterjee, and D. Khovalyg, Dynamic personalized human body energy expenditure: Prediction using time series forecasting LSTM models. Biomedical Signal Processing and Control, 2024. 87: p. 105381.
[12] Maheri, M., et al., Machine learning to assess CO2 adsorption by biomass waste. Journal of CO2 Utilization, 2023. 76: p. 102590.

Disclaimer/Publisher's Note: The perspectives, opinions, and data shared in all publications are the sole responsibility of the individual authors and contributors, and do not necessarily reflect the views of Sciences Force or the editorial team. Sciences Force and the editorial team disclaim any liability for potential harm to individuals or property resulting from the ideas, methods, instructions, or products referenced in the content.