Analysis of Learning Algorithms for Predicting Carbon Emissions of Light-Duty Vehicles

Rashmi B. Kale[1], Nuzhat Faiz Shaikh[2]
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering (SPPU), Pune, Pune, India[1]
Department of Computer Engineering, Wadia College of Engineering, Pune, Pune, India[2]

***Abstract*—This research presents a comparative analysis of different learning methods developed for the prediction of carbon emissions from light-duty vehicles. With the growing concern over environmental sustainability, accurate prediction of carbon emissions is vital for developing effective mitigation strategies. The work assesses the performance of various algorithms trained on vehicle-specific data attributes to predict the emission patterns of a fuel type of different light duty models. This work uses two real-time petrol and diesel datasets collected by CariQ app and device. Canada government dataset is also used from the online repository for prediction of the vehicle emission. The evaluation is based on their predictive accuracy. The findings reveal insights into the effectiveness of different learning techniques in accurately estimating carbon emissions from vehicles, providing valuable guidance for policymakers and researchers in the field of environmental sustainability and transportation planning.**

Keywords—Carbon emission; machine learning algorithms; CariQ carbon emission dataset; An Air Quality Index (AQI)

I. Introduction

As we all know the major threat in front of the world is pollution in the present days. Emission is a problem faced by many countries, and the transportation industry which is one of the biggest sources of carbon allowed to be released into the air [1]. Comprehending and reducing the harmful impacts of vehicle emissions on the atmosphere, living organisms, and the environment has become crucial. Vehicle emissions are a complex mixture of gases and particulate matter, carbon monoxide (CO), nitrogen oxides (NOx), hydrocarbons (HC), and particulate matter (PM). Various detection methods have been employed to monitor and measure these emissions, ranging from traditional tailpipe testing to advanced sensor-based technologies [2] [3]. Government agencies mostly used the term AQI to check the air quality of any area. It is decided through the range of pollution values decided for an AQI [4] [5]. These defined range decide, how is the air quality index of any particular area. Delhi is always with sequious AQI in India and someday pollution reached to its top enough like Most of time and many news flashed about it. So, how important it is to understand the hazards of the emissions and their effects on the living creature mostly on humans? Government needs to put some strict policies about the carbon emission.

The carbon trading system [6], was adopted by many nations to cut greenhouse gas emissions. This market-based strategy served to mitigate the effects of global warming, which improved the air quality in an environment. It is employed to lower the atmospheric concentration of CO2. Carbon credit provides financial incentives to businesses that have emitted less carbon, these businesses are granted a permit of carbon credit. Businesses that generate more carbon emissions need to credits from businesses that have produced lesser emissions. The transport industry significantly contributes to greenhouse gas emissions. To mitigate this impact, the government must establish policies, rules, and regulations specifically targeting emissions from individual vehicles. Here, this research work focuses on the insights of the emissions on light-duty vehicles of diesel and petrol.

The rising concern about global climate change and environmental sustainability has made the theoretical analysis and prediction of CO2 emissions vital. Having an accurate model to forecast CO2 emissions could help policymakers, researchers and industrials to figure out how to better plan to mitigate their environmental impact. There are common methods from classical statistics that might not be able to capture all patterns since the emission of CO2 is a little more complex in a non-linear way - that is why one must look not only to classical statistical methods but also to new computational analysis. This research introduces a hybrid approach [20] that combines moving average smoothing with Long Short-Term Memory (LSTM) neural networks to model and forecast CO2 emissions per one hundred kilometres. Raw CO2 emission data is passed through a moving average filter to suppress noise and reveal low-frequency trends. A model, LSTM, known for its capability to handle time series data is then trained on that smoothed data to predict the future CO2 emissions. A type of recurrent neural network (RNN), LSTM networks are well-suited to this task thanks to their ability to maintain long-term dependencies and deal with the sequential nature of time series data. The proposed methodology tries to achieve more precise CO2 emission forecasts by exploiting these properties. Here, the study first preprocess the raw CO2 emission data by smoothing the data with a moving average filter with a window size of two, and then create two time series for modeling. Now we prepare a dataset suitable for the LSTM model by creating a series of time-lagged input-output pairs in the data. Finally, a dataset is prepared to train a computer using an LSTM model. Predictions are then made, and the Mean Squared Error is calculated.

The findings in this study indicate that moving average filtering combined with the use of LSTM networks could successfully predict CO2 emissions. This approach provides convenient time series analysis tools in environmental studies and increases the predictive accuracy. This work has important implications which present a novel method in which to be able to forecast for any domain where very accurate time series forecasts are required. This section provides information on the

Emissions of Light Duty Vehicles prediction system. The second part includes the literature survey of the system and in addition the Air Quality Index chart (AQI). The third section discussed the proposed work and in the last section, all the results are locally minimized.

II. Literature Survey

The swift growth of urbanization and industrialization has caused a notable surge in global air pollution levels. Among the various pollutants, carbon dioxide emissions are particularly high. This study [7], [8], investigates the prediction of vehicle carbon emissions on metropolitan road networks through the use of sensor networks and real-time data analytics. It explores enhancing the accuracy of emission forecasts by utilizing prediction algorithms, data fusion techniques, and optimal sensor placement. Navigation systems [9], have gotten more and more data-driven in recent years due to the emergence of mobile computing devices and the widespread use of GPS technology. This work presented a navigation system based on traffic incident notification in 2012. Numerous base stations, communication networks, and sensors make up the system. It decides feasible routes to drivers through navigation paths and reduce the vehicular emissions. This study in [10], does a thorough examination of vehicle carbon emissions across time, focusing on temporal dynamics to detect seasonal fluctuations, peak emission periods, and long-term trends. In order to anticipate future emission trends and assist in the formulation of policy, it creates predictive models. In order to anticipate carbon emissions, this study [11], looks into the integration of vehicle telematics data, such as GPS trajectories, engine performance indicators, and fuel consumption rates. It looks at how telematics-enabled technologies can improve fleet operations and lessen their negative effects on the environment. In order to forecast vehicle carbon emissions under different policy interventions, such as fuel efficiency standards, emission laws, and investments in transportation infrastructure, this study [12], uses scenario analysis methodologies. It assesses how well various policy scenarios work to meet emission reduction goals. This study in [13], investigates predictive maintenance techniques with an emphasis on fleet management tactics in order to maximize vehicle performance and minimize carbon emissions. In order to improve fleet sustainability, it looks at how to integrate condition monitoring, remote diagnostics, and predictive analytics. Government organizations mostly employed the term "AQI" to assess the air quality of any area. The index values and AQI [14], [15], color shown in the Table I determine the range of the air quality, based on these concern levels are decided. Most of the metropolitan cities falls under maroon to a red zone that is very unhealthy and hazardous for living creature.

TABLE I. Air Quality Index Chart

AIR QUALITY INDEX CHART			
AQI Color	***Levels of Concern***	***Index Value***	***Quality of Air and effects on human***
Green	Good	0 to 50	Efficient air pollution does not present significant risks.
Yellow	Moderate	51 to 100	Acceptable, but for some who are sensitive to air pollution, may have a risk.
Orange	Unhealthy for Sensitive Groups	101 to 150	Poor, sensitive people may experience health effects.
Red	Unhealthy	151 to 200	Very Poor, sensitive people may experience more serious health effects.
Purple	Very Unhealthy	201 to 300	Extremely poor
Maroon	Hazardous	301 and higher	Extremely worst

III. Proposed Work

A. Data Collection

This work is focusing on the automobile sector and the emission patterns of different vehicle models, fuel types of light duty vehicles. This research utilized a mobile application called CariQ to collect real-time carbon emission data from vehicles operating within Pune City for nearly two years. The resulting dataset serves as the foundation for this project [16], [17], [18], [19]). CariQ Mobile app screen shot is shown in Fig. 1.

In Fig. 2, it shows mixed dataset of all petrol and diesel fuel type also created. Fig. 2 shows the screenshot of some samples of dataset created through CariQ device and mobile app.

B. Proposed Methodology

This study's methodology encompasses a sequence of steps designed to process and model CO2 emission data, employing both a moving average filter and an LSTM (Long Short-Term Memory) neural network.

1) Data Preparation

a) Initial Data

- Input Data: *per_hundred_KM_CO2_emission* is an array representing CO2 emissions per 100 km for a series of observations.

b) Smoothing Data

- Moving Average Filter: Apply a moving average filter with a window size of 2 to smooth the data. This reduces noise and helps in trend detection.

2) Dataset Creation for LSTM

a) Sequence Generation

- Look-back Window: Define a look-back window (*look_back = 1*) to create input-output pairs for the LSTM model. This means the model uses the value at time t to predict the value at time $t+1$.
- Function create_dataset: This function generates sequences of inputs and corresponding outputs from the smoothed data.

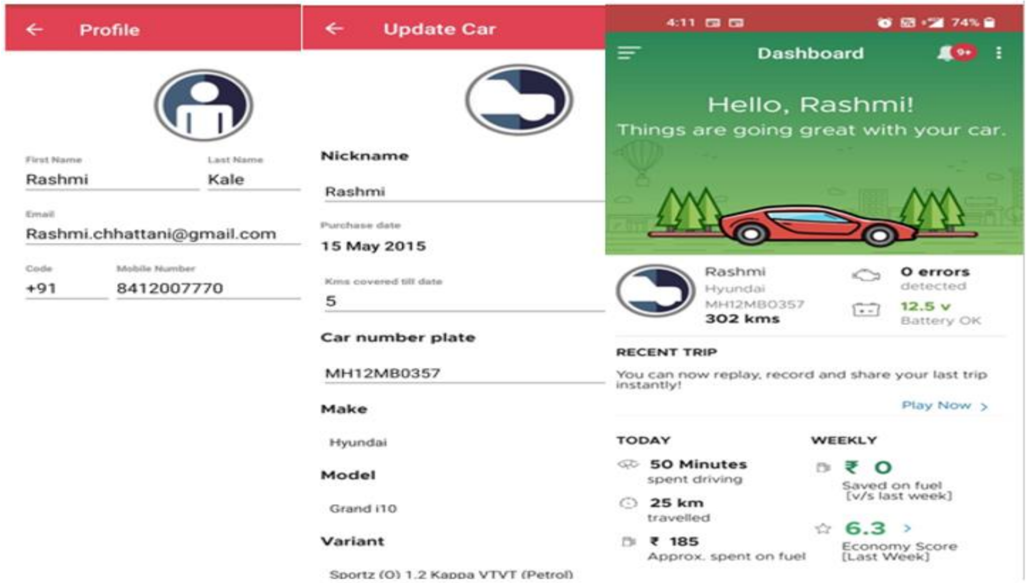

Fig. 1. Screen shot of CariQ app for creating the dataset.

SR. NO.	VEHICLE TYPE	COMPANY	MODEL	ENGINE SIZE	CYLINDERS	VEHICLE NO.	DATE OF PURCHASE	FUEL TYPE	AMOUNT SPEND ON FUEL	FUEL CONSUMPTION (Lit.)	KM TRAVELLED	CO2 EMISSION	Petrol PRICE	EMISSION FACTOR
1	CAR	HYUNDAI	i20	1.2	4	MH 05 CM-5993	12/12/2015	Petrol	185	1.73	25.00	4.00	103	2.296
2	CAR	Maruti	Swift Dzire	1.2	4	MH 28 V-3753	15/03/2012	Diesel	189	2.1	32.0	5.5	92.03	2.68
3	CAR	HYUNDAI	GRAND i10	1.2	4	MH 12 MB-0357	15/5/2015	Petrol	1160	10.84	148.00	25.12	104	2.296
4	CAR	Volkswagen	Jetta	2.0	4	MH 17 AZ-2261	1/9/2014	Diesel	162	1.7	27.0	4.6	95.37	2.68
5	CAR	TATA	Zest	1.1	4	MH 12 NJ-1684	12/9/2016	Diesel	101	1.0	16.8	2.8	96.35	2.68
6	CAR	HYUNDAI	GRAND i10	1.2	4	MH 12 MB-0357	15/5/2015	Petrol	1160	10.84	148.00	25.12	104	2.296

Fig. 2. Screen shot of some samples of created dataset.

3) LSTM Model Training

a) Data Reshaping

- Reshape Input Data: Reshape the input data to the required shape for LSTM, which is [*samples, time steps, features*].

b) Model Definition

- LSTM Model: Define an LSTM model with one hidden layer containing four units.
- Dense Layer: Add a Dense layer with one unit to produce the output.

c) Model Compilation and the Training

- Compile Model: Compile the model with Mean Squared Error (MSE) as the loss function and the Adam optimizer.
- Fit Model: Train the model on the training data for 100 epochs with a batch size of 1.

4) Making Predictions

a) Predictions on Training Data

- Train Predictions: Use the trained model to predict values on the training data.

b) Future Predictions

- Future Values: Predict future values for an extended range using the model.

5) Model Evaluation

a) Calculate MSE

- Mean Squared Error: Evaluate the model's performance by calculating the Mean Squared Error between the actual smoothed data and the predicted values.

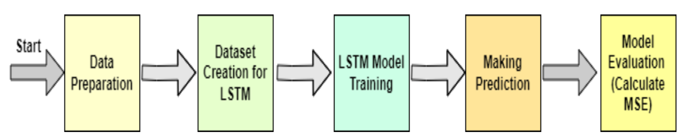

Fig. 3. Proposed methodology for predicting carbon emissions.

This methodology involves data smoothing using a moving average, preparing data sequences for LSTM input, defining and training an LSTM model, making predictions, and evaluating the model's performance using Mean Squared Error as shown in Fig. 3. This approach leverages deep learning to model time series data effectively.

IV. Results And Discussion

Around 175 Diesel and 125 Petrol Vehicles are considered to create Carbon Emission dataset. Fig. 4 illustrates the number of Diesel and Petrol vehicles included in the dataset. This dataset combines various types of light-duty vehicles and fuel types for assessment.

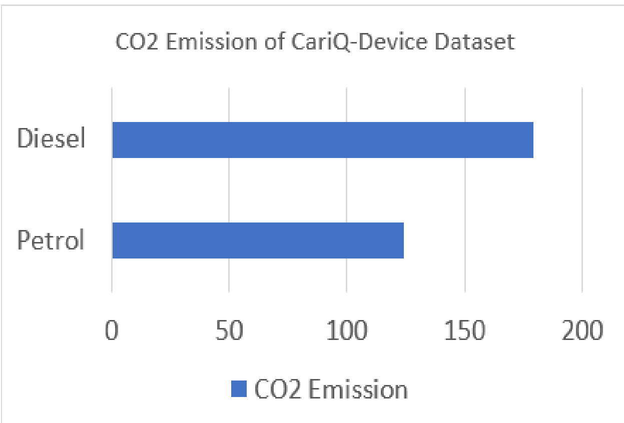

Fig. 4. Number of petrol and diesel vehicles.

Different machine learning algorithms were applied to two different datasets of petrol and diesel vehicles to find out the emission patterns. Fig. 5 shows a scatterplot of carbon emission in gm/km through the diesel vehicle.

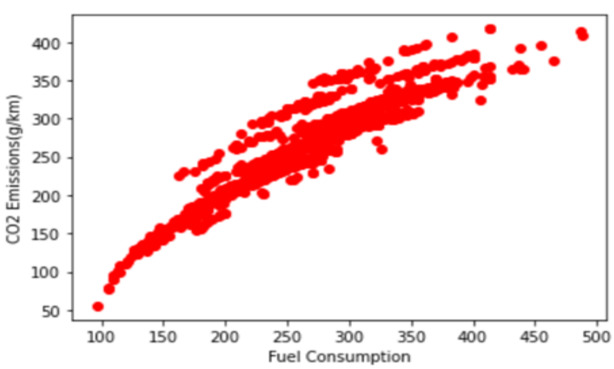

Fig. 5. Diesel vehicle carbon emission.

Fig. 6 shows scatterplot of carbon emissions in gm/Km through a Petrol vehicle. From both the petrol and diesel dataset results, it has been observed that a diesel vehicle produces more carbon emission than the petrol vehicle.

Fig. 7 shows boxplot of Carbon emissions of both petrol and diesel vehicles.

Now, after statistical analysis step, Machine learning algorithms are applied to the dataset, Based on this mean square error for each algorithm is shown in Fig. 7.

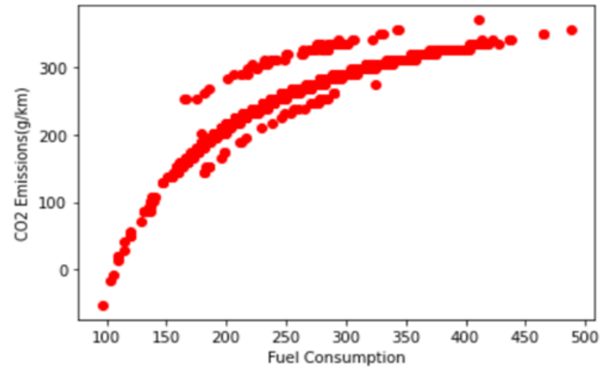

Fig. 6. Petrol vehicle carbon emission.

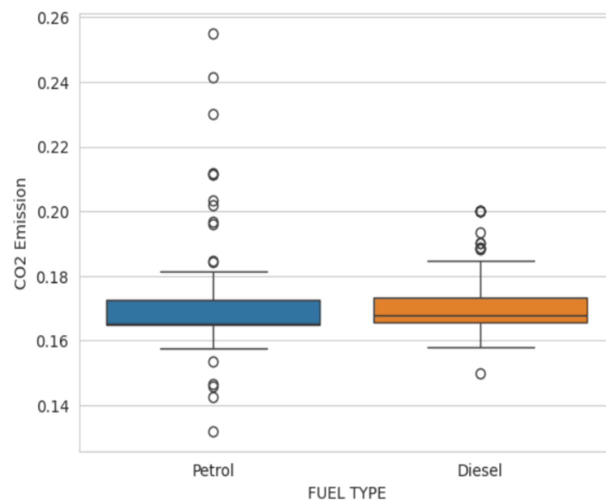

Fig. 7. Boxplot of CO2 Emissions by petrol and diesel vehicles.

Fig. 8 and Table II show MSE values got from the different machine learning algorithms. Proposed system algorithm applied to the dataset gives minimum MSE value than other algorithms. Proposed system algorithm is the advanced LSTM algorithm.

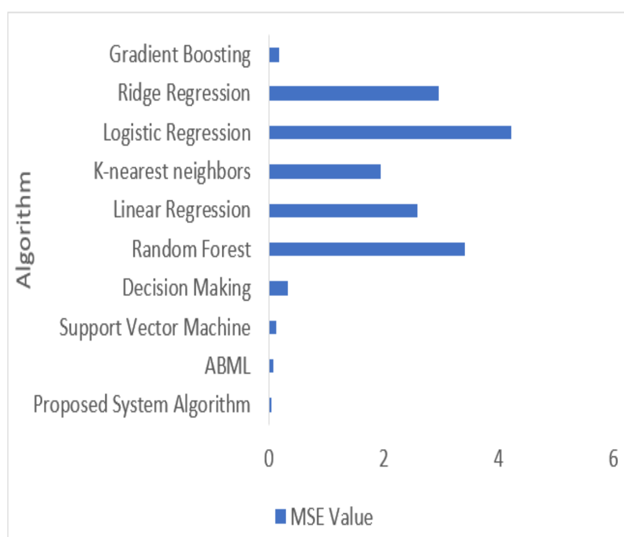

Fig. 8. MSE values for a different machine learning algorithms.

TABLE II. MSE VALUES FOR DIFFERENT MACHINE LEARNING ALGORITHMS

Learning Algorithms	***MSE***
Gradient Boosting	0.179
Ridge Regression	2.95
Logistic Regression	4.218
K- nearest neighbors	1.942
Linear Regression	2.591
Random Forest	3.414
Decision Making	0.328
Support Vector Machine	0.128
ABML	0.083
Advanced LSTM (Ours)	0.047

Another dataset used for vehicular carbon emission is Canada government dataset. Vehicle carbon emission data assessment provides comprehensive information on emissions produced by various types of vehicles and a different type of fuel. Fig. 9 shows CO_2 emission by five different fuel types' natural gas, diesel, regular gasoline premium gasoline and ethanol (E85). Through this dataset assessment, it is found that ethanol vehicle emission more than other fuel types.

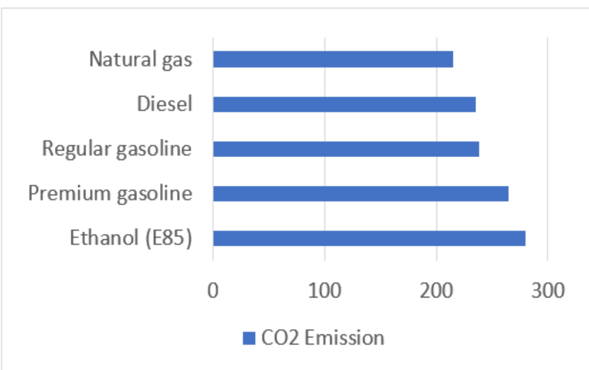

Fig. 9. Carbon emission for different fuel types of Canada government dataset.

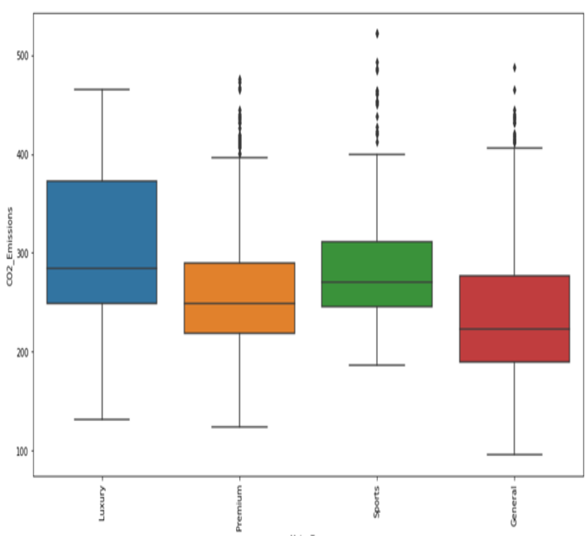

Fig. 10. Boxplot of carbon emission by various light duty vehicle.

Fig. 10 shows the boxplot of CO2 emission of vehicles for different light duty vehicle categories. From this analysis it has been analysed that luxury vehicles produces more emissions than other category vehicles.

V. CONCLUSION

This study emphasizes the significance of predicting carbon emissions from light duty vehicles amidst growing concerns for environmental sustainability. Through this comparative evaluation of machine learning-based approaches, it assesses the performance of various algorithms trained on a vehicle-specific data attributes. Leveraging real-time datasets from petrol and diesel vehicles collected via the CariQ app and device, alongside Canada government data. Data analysis focused on predictive accuracy. The various types of vehicle data studied and concluded that a luxury vehicle and diesel vehicles emit more carbon in the environment. The findings shed light on the effectiveness of different machine learning techniques in estimating carbon emissions, offering valuable insights for policymakers and researchers engaged in environmental sustainability and transportation planning. By adopting and promoting multi-modal transportation by investing in and encouraging public transit infrastructure and active transportation. By putting these suggestions into practice, academics and policymakers may strive toward creating more resilient and sustainable transportation networks that lessen their negative effects on the environment and enhance everyone's quality of life.

REFERENCES

[1] Khurana, S., Saxena, S., Jain, S., & Dixit, A. (2020). Predictive modeling of engine emissions using machine learning: A review. Materials Today: Proceedings. doi:10.1016/j.matpr.2020.07.204.

[2] Li, Q., Qiao, F., & Yu, L. (2017). A Machine Learning Approach for Light-Duty Vehicle Idling Emission Estimation Based on Real Driving and Environmental Information. Environment Pollution and Climate Change, 01(01). doi:10.4172/2573-458x.1000106.

[3] Natarajan, Yuvaraj & Wadhwa, Gitanjali & Ramasamy, Sri & Paul, Anand. (2023). Forecasting Carbon Dioxide Emissions of Light-Duty Vehicles with Different Machine Learning Algorithms. Electronics. 12. 2288. 10.3390/electronics12102288.

[4] Peng, T.; Yang, X.; Xu, Z.; Liang, Y. Constructing an Environmental Friendly Low-Carbon-Emission Intelligent Transportation System Based on Big Data and Machine Learning Methods. Sustainability 2020, 12, 8118. https://doi.org/10.3390/su12198118.

[5] Esther Pushpam, V. S., Kavitha, N. S., & karthik, A. G. (2019). IoT Enabled Machine Learning for Vehicular Air Pollution Monitoring. 2019 International Conference on Computer Communication and Informatics (ICCCI). doi:10.1109/iccci.2019.8822001.

[6] Zhao, Dengfeng & Li, Haiyang & Hou, Junjian & Gong, Pengliang & Zhong, Yudong & He, Wenbin & Fu, Zhijun. (2023). A Review of the Data-Driven Prediction Method of Vehicle Fuel Consumption. Energies. 16. 5258. 10.3390/en16145258.

[7] Y. Kai, J. Shuai, H. Chunxuan, Z. Zheng and L. Tianran, "Analysis on the emission reduction benefits of electric vehicle replacing fuel vehicle," 2021 IEEE Sustainable Power and Energy Conference (iSPEC), Nanjing, China, 2021, pp. 3396-3402, doi: 10.1109/iSPEC53008.2021.9735774.

[8] L. Wang, "Carbon emission trading and technology innovation for low-carbon emission," 2011 International Conference on Electrical and Control Engineering, Yichang, China, 2011, pp. 1472-1477, doi: 10.1109/ICECENG.2011.6058285.

[9] Sannigrahi, S., Pilla, F., Basu, B., Basu, A. S., Sarkar, K., Chakraborti, S Roy, P. S. (2020). Examining the effects of forest fire on terrestrial carbon emission and ecosystem production in India using remote sensing

approaches. Science of The Total Environment, 138331. doi:10.1016/j.scitotenv.2020.1383.

[10] Rebolledo-Leiva, R., Angulo-Meza, L., Iriarte, A., & González-Araya, M. C. (2017). Joint carbon footprint assessment and data envelopment analysis for the reduction of greenhouse gas emissions in agriculture production. Science of The Total Environment, 593-594, 36–46. doi:10.1016/j.scitotenv.2017.03.1.

[11] Lin, X., Zhu, X., Han, Y., Geng, Z., & Liu, L. (2020). Economy and carbon dioxide emissions effects of energy structures in the world: Evidence based on SBM-DEA model. Science of The Total Environment, 138947. doi:10.1016/j.scitotenv.2020.1389.

[12] Suzanne Greene a, Haiying Jia b, Gabriela Rubio-Domingo c. (2020). Well-to-tank carbon emissions from crude oil maritime transportation. Transportation Research, 102587. sci-hub.se/10.1016/j.trd.2020.102587.

[13] Xiaojun Ma, Changxin Wang, Biying Dong, Guocui Gu, Ruimin Chen, Yifan Li, Hongfei Zou, Wenfeng Zhang, Qiunan Li (2019). Carbon emissions from energy consumption in China: Its measurement and driving factors. Science of The Total Environment, Volume 648, 15 January 2019, Pages 1411-1420.

[14] Sun, S., Li, L., Wu, Z., Gautam, A., Li, J., & Zhao, W. (2020). Variation of industrial air pollution emissions based on VIIRS thermal anomaly data. Atmospheric Research, 105021. doi:10.1016/j.atmosres .2020.10502.

[15] Luo, L., & Chen, Y. (2020). Carbon Emission Energy Management Analysis of LCA-Based Fabricated Building Construction. Sustainable Computing: Informatics and Systems, 100405. doi:10.1016/ j.suscom.2020.100405.

[16] R. Kale, Nuzhat F. Shaikh (2021). Forecasting Models for Carbon Emission: A survey and Discussion. International Journal of Mechanical Engineering. Vol. 6 No. 3 December, 2021.

[17] R. Kale, Nuzhat F. Shaikh, K. Wankhade (2023). Survey Paper on Smart Automation System for Calculating the Vehicular Carbon Emission using gadget CariQ and its Applications, International conference on Recent Trends in Science, Technology, Engineering & Mathematics (MICRT-STEM 2023), February 24, 2023. ISBN-13: 978-93-5967-432-2.

[18] R. Kale, Nuzhat F. Shaikh (2024). IoT-Enabled Forecasting of Vehicle based Carbon Emissions and Smart Fuel Management: Optimizing Efficiency and Performance. International Journal of Intelligent Systems and Applications in Engineering (IJISAE), February 24, 2024, 12(19s), 161–167, ISSN:2147-6799214.

[19] R. Kale, Nuzhat F. Shaikh (2024). Algorithmic Modeling for Predicting Carbon Emissions in an Individual Vehicles: A Machine Learning and Deep Learning Approach. International Journal of Intelligent Systems and Applications in Engineering (IJISAE)-unpublished.

[20] Ganesh D. Jadhav, Sachin D. Babar and Parikshit N. Mahalle, “Hybrid Approach for Enhanced Depression Detection using Learning Techniques” International Journal of Advanced Computer Science and Applications(IJACSA), 15(4), 2024.