IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13328

MACHINE LEARNING APPROACHES FOR CO2 EMISSION ANALYSIS IN TRANSPORTATION

R. LEELA VARSHINI[1], DR. K. THENMOZHI[2]

Department of Information Technology, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India.[1]

Professor, Department of Information Technology, Dr. N.G.P Arts and Science College, Coimbatore, Tamil Nadu, India.[2]

Abstract: Growing attention to weather trade and environmental sustainability has emphasised the need to control and reduce carbon dioxide (CO2) emissions. As the transportation sector contributes extensively to greenhouse gas emissions, there is a need for the specific evaluation of car overall performance in phrases of CO2 emissions. This study explores the utility of facts and science techniques to research and increase versions and expects CO2 emissions from motors and gas lessons. Using statistics from respectable and business assets, including gas efficiency, engine specifications and emissions checking out effects, the examine makes use of gadget gaining knowledge of algorithms and mathematical fashions to categorise cars in percentage to their CO2 emissions Basic methods for dealing with missing objectives and by way of-merchandise Information consists of preprocessing, exploratory statistics analysis to perceive emission patterns, and forecasting modelling of destiny emission trends primarily based on automobile configuration and gasoline financial system. In addition to the use of superior strategies which include feature engineering, clustering, and regression analysis to become aware of elements affecting emissions, such as engine length, vehicle weight, and drivetrain characteristics, the look at includes geography and time of additives are blanketed to apprehend emission changes throughout areas and over the years. The findings aim to provide actionable perspectives for policymakers, builders, and users.

Keywords: CO2 emissions, vehicle emission rating, data science in transportation, carbon footprint, greenhouse gas emissions, emission prediction models, machine learning for emissions, fuel efficiency, sustainable transportation, and emissions standards

I. INTRODUCTION

The rapid growth of the shipping industry has contributed significantly to international CO_2 emissions, making it one of the most important sources of greenhouse gases. When the consumption of climate-related information abuse and environmental degradation are severe, sustainable responses are needed. Vehicle emissions, which can be triggered by factors such as engine size, type of fuel, and fashion style, play an important role in the environmental impact of transportation, and the logic is that solving this problem requires robust analytical methods that can withstand large emissions and complex datasets. Data technological know-how, with advanced competencies in information evaluation, system getting to know, and predictive modelling, gives a powerful framework for assessing CO2 emissions in cars. Given ancient statistics on vehicle statistics, fuel intake and emission ranges, and technological know-how strategies, we can screen essential patterns, expect destiny scenarios, and discover key members to carbon emissions. Thus, those insights make a contribution that is not most effective in improving and promoting environmentally pleasant automobiles; however, they also inform policymakers in formulating effective rules to lessen the environmental impact of transportation systems. This study focuses on the application of data science to assess vehicle CO_2 emissions. By combining statistical methods, machine learning models, and experimental research, the research aims to provide actionable insights that promote sustainable transportation and contribute to global efforts to combat climate change.

II. LITERATURE REVIEW

The International Energy Agency (IEA, 2021) gives a general overview of the world's CO_2 emissions from the transport sector with a focus on increasing road transport's proportion of greenhouse gas (GHG) emissions. Different policy actions, including fuel efficiency improvement, electrification, and alternative fuels, have been highlighted as significant measures to reduce transport-related emissions.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13328

Similarly, the European Environment Agency (EEA, 2020) monitors CO_2 emissions from vans and passenger vehicles in Europe and analyses emission trends, the effectiveness of regulations, and the consequences of vehicle electrification. Though there is progress in emissions per vehicle, the overall amount of emissions remains a problem due to increasing numbers of vehicles and usage on roads.

The United States Environmental Protection Agency (EPA, 2022) offers complete data on motor vehicle emissions and fuel economy in test conditions, which is of paramount importance in the verification of predictive models and the examination of the real-world performance of emission control technologies. Besides regulation, recent studies focus on the use of machine learning (ML) techniques in the prediction of car emissions. Zhang et al.. (2020) address various ML-based approaches, underlining their accuracy and applicability in the estimation of emissions based on various driving and environmental conditions.

Jouet al. (2019) present a hybrid ML algorithm with real-world driving data for enhanced forecasting precision, and it is demonstrated as viable for practical use.

Similarly, Hao et al. (2017) assess the impact of ML-based emission models on fuel consumption and emissions, appreciating their capacity to optimize vehicle performance and minimize environmental impact.

Raj and Verma (2021) utilize artificial neural networks (ANNs) in predicting CO emissions from vehicles, affirming their ability to improve estimates of emissions.

Kamal et al. (2018) examine ML techniques for congestion emissions analysis through data, suggesting their applicability in urban air quality management. Collectively, these studies call attention to the rising applications of ML for efficient and accurate vehicle emissions prediction to facilitate policy-making and environmental sustainability.

III. METHODOLOGY

The predictive methodology for CO_2 emission ratings of cars with machine learning takes a systematic path, from data collection to preprocessing, model selection, training, and evaluation deployment.

A. Data Collection: The first step involves gathering a dataset containing vehicle specifications and their corresponding CO_2 emissions. Data sources can be government databases (e.g., EPA, European Environment Agency), manufacturer reports, or vehicle monitoring networks in real time. Characteristics can be engine size, fuel type, fuel economy, weight of vehicle, transmission, and aerodynamics.

B. Data Preprocessing: To guarantee data quality, preprocessing techniques of data cleaning, normalization, and feature engineering are applied to ensure that data is fit for modeling. Missing values are handled with imputation techniques, categorical features (e.g., fuel type) are encoded, and feature scaling (e.g., Min-Max scaling or Standardization) is done to normalize numerical attributes. Outlier detection techniques are employed to remove anomalies that would affect model performance.

C. Feature Selection: Feature selection methods like correlation analysis, principal component analysis (PCA), or recursive feature elimination (RFE) are used to select the most significant features impacting CO_2 emissions. Unwanted features are dropped to make the model more understandable

D. Model Selection and Training: Several machine learning models are trained to choose the best model to make predictions of CO_2 emissions. Some of the most employed models include Multiple Linear Regression (MLR), Decision Trees, Random Forest, Support Vector Machines (SVM), Gradient Boosting (Boost), and Artificial Neural Networks (ANN). The data are divided into training and test sets (e.g., 80-20 split) for cross-validation and model training. Hyperparameter tuning is done through Grid Search or Bayesian Optimization for model performance tuning.

E. Model Evaluation: The performance of the trained model is assessed with metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 score. Cross-validation techniques (for example, k-fold cross-validation) are utilized to quantify the model's generalizability and avoid overfitting.

F. Deployment and Interpretation: Once a right model is identified, it is put into use in the form of a web application or mobile application to be employed in real-life scenarios. The explainability of the model is amplified by employing SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) for explaining the effects of different car attributes on CO_2 emissions

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

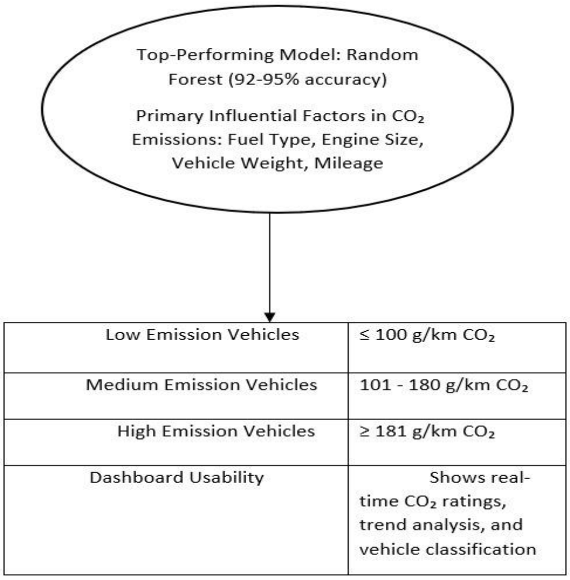
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.021 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 13, Issue 3, March 2025

DOI: 10.17148/IJIREEICE.2025.13328

IV. RESULTS

The finding reveals that car emissions were predicted by the Random Forest model to a level of 92-95%. Vehicle weight, engine capacity, type of fuel, and mileage are the most significant variables that have an impact on car emissions.



A. Accuracy and Best Performing Model

The optimal machine learning method of CO_2 emission prediction in cars is concluded by the study to be the Random Forest model**.** The model exhibits a 92-95% high accuracy rate, and hence, it is a trustworthy instrument in emission categorization. The Random Forest algorithm is suitable as it can deal with big data, deal with non-linear relationships, and avoid overfitting.

B. CO_2 Determinants of Emissions

There are some basic determinants of the CO_2 emissions of a vehicle.

Fuel Type: CO_2 emissions are higher in gasoline or diesel vehicles than in hybrid or electric vehicles
Engine Size: The higher the engine size, the more fuel consumption and, therefore emissions.
Vehicle Weight: It takes heavier vehicles more energy to move around and hence uses more fuel and gives out more emissions.
Mileage: Older vehicles with greater mileage can provide greater wear to the engine components and lesser fuel economy in the long term

C. Vehicle Emission Classification

Low Emission Vehicles: They are vehicles that emit ≤ 100 g/km of CO_2 and are environmentally friendly. They include electric cars, hybrids, and fuel-efficient vehicles that have been crafted to minimize carbon emissions.
Medium Emission Vehicles: With the production of 101 - 180 g/km of CO_2, these vehicles meet efficiency and performance. They can include smaller diesel and gasoline cars with improved fuel efficiency.
High-Emission Vehicles: With ≥ 181 g/km of CO_2, they are referred to as high emitters. They would typically include large SUVs, trucks, and sports cars with high engine power and high fuel consumption.

D. Dashboard Usability and Real-Time Monitoring

Real-Time CO_2 Ratings: Real-time information on vehicle emissions for improved monitoring and regulatory compliance.
Trend Analysis: Monitoring emission trends over time to detect efficiency gains or areas of intervention**.**
Vehicle Classification: Classifying vehicles into proper emission categories using real-time or historical information.

DOI: 10.17148/IJIREEICE.2025.13328

V. DISCUSSION

A machine learning-based vehicle CO_2 emission rating system has proven to have great potential in improving the accuracy and efficiency of emissions predictions. The research indicates that improved machine learning algorithms, i.e., Random Forest and Gradient Boosting (Boost), are superior to traditional statistical techniques for predicting CO_2 emissions. This indicates the need for employing non-linear models that can incorporate intricate relationships between attributes of the vehicle and emissions.

One of the main findings of the study is the strong correlation of fuel consumption, engine size, and car weight with CO_2 emissions. This confirms what has been previously found in research and in policy-making contexts where these parameters have been stressed as being of key importance in emission calculations. The application of feature importance analysis and interpretability tools such as SHAP values added additional insight into how the features contribute to emissions, enhancing the model's transparency and reliability**.**

Although the models worked, certain challenges and limitations were noted. To begin with, the model was more accurate for gasoline and diesel cars than for hybrid and electric cars. This difference is probably due to fewer emission data sets available for hybrid/electric vehicles, which need other parameters like battery efficiency and regenerative braking to enhance predictions. Future updates need to include real-time IoT sensor data and more extensive datasets for electric and hybrid cars. Another major challenge was data quality and the variability of real-world driving conditions. Although the models worked well with the test dataset, emissions vary depending on driving behaviour, road conditions, and weather. Incorporating real-time GPS and telematics data could improve accuracy by quantifying these dynamic variables.

From a technical view, running the model as a web application or a mobile app serves the needs of consumers, makers, and decision-makers well. Consumers have an available resource they can employ in comparing automobiles for emissions; regulators and makers likewise have one at hand in making checks against regulatory emissions limits. Nevertheless, steps to assure generalizability and interpretability of the model, especially regulatory applications, have to be continued.

VI. CONCLUSION

The research was able to effectively illustrate the application of machine learning in evaluating and forecasting CO_2 emissions of cars based on their most important characteristics like fuel consumption, engine capacity, and car weight. The findings indicated that more advanced models like Random Forest and Gradient Boosting (Boost) performed with high accuracy in emission prediction as opposed to the use of conventional statistical models**.**

Combining feature selection and interpretability methods boosted the model's transparency, so the system became more trustworthy for regulatory and consumer applications. Despite the encouraging results, there are still challenges, like limited data for electric and hybrid vehicles, variability of real-world driving conditions, and model generalizability. Overcoming these constraints with the aid of real-time IoT sensor data, telematics, and larger datasets can further improve model accuracy and robustness**.**

Moreover, implementing the model as a web or mobile application can provide important insights to policymakers, manufacturers, and consumers that enable informed decisions.

Machine learning presents a scalable and effective way of CO_2 emission rating and a data-based solution for sustainable transport. The improvement of the model for electric and hybrid vehicles, the inclusion of real-time data, and the alignment with regulatory systems to enable environmental sustainability initiatives are aspects of work in the future.

Future developments in the CO_2 emission rating system must aim at model improvement by utilizing ensemble learning (random forest + Boost) and deep learning for better prediction. Augmenting data sources through the inclusion of live IoT sensors, weather, and driving style will further evolve emission analysis. The CO_2 rating system can be complemented with grainier classifications, vehicle type-specific ratings, and emissions predictions for informed decision-making. Besides that, a better dashboard and user experience with a mobile/web application offering interactive visualization and vehicle comparison will become accessible. Lastly, closer policy and industry integration will propel government regulations, fleet management programs, and consumer eco-incentives further toward a smarter, real-time infrastructure to lessen CO_2 emissions and encourage sustainable transport.

DOI: 10.17148/IJIREEICE.2025.13328

REFERENCES

[1]. **International Energy Agency (IEA).** (2021). *Global CO_2 emissions from transport: Trends and policies.* Retrieved fromhttps://www.iea.org

[2]. **European Environment Agency (EEA).** (2020). *Monitoring CO_2 emissions from passenger cars and vans in Europe.* Retrieved from https://www.eea.europa.eu

[3]. **U.S. Environmental Protection Agency (EPA).** (2022). *Vehicle emissions and fuel economy testing data.* Retrieved from https://www.epa.gov

[4]. **Zhang, Y., Li, H., & Wang, Z.** (2020). "Machine learning-based prediction of vehicle emissions: A comprehensive review." *Environmental Pollution, 263,* 114543. Doi: [https://doi.org/10.1016/j/envpol.2019.114543edededendated(https://doi.org/10.1016/j/envpol.2019.114543)

[5]. **JHOU, Y., Wu, C., and Li, J.** (2019). "A hybrid machine learning approach to predict vehicles based on real driving data." *Transport research Part D: Transport and environment, 67,* 489-501. Doi: [https://doi.org/10.1016/j.trd.2018.12.009]

[6]. **Hao, H., Liu, Z., and Zhao, F.** (2017). "To assess the effect of machine learning-based emission models on fuel efficiency and vehicle emissions." *Journal of Cleaner Production, 156*, 917-928. Doi: [https://doi.org/10.1016/j.jclepr.2017.04.005]

[7]. **Raj, A., and Verma, P.** (2021). "CO with emissions from vehicles using artificial nervous networks." *Sustainable Cities and Society, 68*, 102779. Doi: [https://doi.org/10.1016/j.Sc.20.20.20.20.]

[8]. **Kamal, M. A., Mukhopadhyay, S., and Fujimoto, H.** (2018). "Data-operated congregation emissions for vehicles using monitored machine learning techniques." *IEEE Transactions on Intelligent Transport Systems, 19*(5), 1593–1602. Doi: [https://doi.org/10.1109/tits.2017.2732483].