

Machine Learning Prediction of CO₂ Emissions from Light-Duty Vehicles in Canada

Barka Satya

Department of Informatics Engineering, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
barka.satya@amikom.ac.id (corresponding author)

Muhammad Daffa Miquilla

Department of Informatics Engineering, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
miquilla@students.amikom.ac.id

Afrig Aminuddin

Department of Information Systems, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
afrig@amikom.ac.id

Ahmad Naufal Labiib Nabhaan

Department of Computer Engineering, Faculty of Computer Science, Universitas Amikom Yogyakarta, Sleman, Indonesia
ahmad.nabhaan@students.amikom.ac.id

Mohammad Badrul Alam Miah

Department of Information and Communication Technology, Mawlana Bhashani Science and Technology University, Tangail, Bangladesh
badrul.ict@mbstu.ac.bd

Hesmeralda Rojas Enriquez

Escuela Academico Profesional de Ingenieria Informatica y Sistemas, Facultad de Ingenieria, Universidad Nacional Micaela Bastidas de Apurimac, Abancay, Peru
hrojas@unamba.edu.pe

Received: 17 October 2025 | Revised: 25 November 2025, 10 December 2025, and 16 December 2025 | Accepted: 17 December 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.15599>

ABSTRACT

The accurate prediction of CO₂ emissions from light-duty vehicles is crucial for effective environmental regulation and policy development. Addressing the limitations of previous studies that often rely on single-model approaches, the present research establishes a novel and rigorous performance benchmark by systematically evaluating seven distinct Machine Learning (ML) architectures—ranging from linear baselines to deep neural networks—to identify the optimal predictive framework for the Canadian context. The study evaluated the performance of Linear Regression, Ridge, Random Forest, Gradient Boosting, Support Vector Regression (SVR), K-Nearest Neighbors (KNN), and a Neural Network by utilizing a public dataset of 7,385 vehicles. Following systematic hyperparameter tuning, the Random Forest model demonstrated superior performance, achieving an R^2 of 0.9982 and a Root Mean Square Error (RMSE) of 2.49 g/km on the test set. Feature importance analysis confirmed that combined fuel consumption is the most dominant predictor of CO₂ emissions. This study establishes a new performance benchmark for CO₂ emission modeling in the Canadian context. The former offers a robust, data-driven tool for regulators and the automotive industry to support emission reduction strategies.

Keywords-machine learning; predictive modeling; vehicle emissions; regression analysis; CO₂ emissions

I. INTRODUCTION

The escalating climate crisis, driven by anthropogenic greenhouse gas emissions, demands swift and decisive action. Carbon dioxide (CO₂) is the primary contributor to global warming, and its mitigation is central to international environmental agreements and the United Nations' Sustainable Development Goals (SDGs) [1]. Specifically, efforts to reduce CO₂ align with SDG 13 (Climate Action) and are critical for achieving SDG 11 (Sustainable Cities and Communities) by fostering healthier urban environments [2]. The transportation sector, a major source of global CO₂ emissions, has thus become a key focus for policy intervention and technological innovation [3]. Accurately modeling and predicting vehicle emissions is significant in developing effective strategies to decarbonize this sector.

Canada has established ambitious environmental targets within this global landscape, including a commitment to having achieved net-zero emissions by 2050 [4]. With the transportation sector accounting for approximately 25% of the nation's total greenhouse gas emissions, light-duty vehicles represent a major challenge and a critical area for regulatory focus. This challenge is not unique to Canada; similar trends are observed globally. For instance, in the European Union, the transport sector accounts for nearly 30% of the total CO₂ emissions, with 72% of emissions originating from road transport [5]. This alignment highlights the universality of the problem and the urgent need for robust predictive models across different jurisdictions. To guide its climate strategy, the Canadian government relies on data-driven policies that require precise monitoring and verification of vehicle emissions [6]. This necessity has spurred interest in advanced analytical techniques that can model the complex relationships between vehicle specifications and their resulting environmental impact.

ML application is a powerful approach for this task, frequently outperforming traditional statistical methods [7]. Previous research has demonstrated the efficacy of various ML algorithms—such as Random Forest and Gradient Boosting—in predicting vehicle emissions in diverse international contexts, including China and the United States [8]. While some studies have explored ML applications for Canadian data, a comprehensive and systematic comparison across a broad spectrum of models, from simple linear regressions to more complex neural networks, remains an underexplored area [9]. A critical gap exists in establishing a methodologically rigorous benchmark that identifies the optimal predictive model specifically tuned for the nuances of the Canadian vehicle fleet.

The present study proposes a robust comparative framework to identify the most effective model for predicting CO₂ emissions from light-duty vehicles in Canada, and thus address this gap. The former systematically evaluates seven distinct ML algorithms: Linear Regression, Ridge Regression, Random Forest, Gradient Boosting, SVR, KNN, and a Neural Network. These specific algorithms were selected to cover a broad spectrum of model complexity, ranging from interpretable linear baselines to advanced ensemble and deep learning architectures. This diversity allows for a rigorous assessment of predictive performance relative to computational

cost. Furthermore, this data-driven approach established physical emission models, such as the COPERT methodology [10], by providing granular, vehicle-specific predictions based on real-world configuration data rather than aggregate emission factors.

The primary objectives of the present study are twofold: first, to determine the model with the highest predictive accuracy through rigorous hyperparameter tuning and evaluation using the R^2 and RMSE metrics; and second, to identify the most influential vehicle characteristics driving CO₂ emissions through a detailed feature importance analysis. Furthermore, the selection of Canada as the case study is primarily driven by the high quality and comprehensive nature of its open government data. While the current work is based in regions where such granular vehicle data are currently being developed, the Canadian dataset provides a reliable environment for validating complex ML architectures. This ensures that the proposed framework is methodologically sound before being adapted to other regions with challenging data availability.

II. RELATED WORKS

Significant research has focused on forecasting aggregate national and sectoral CO₂ emissions using advanced time-series and hybrid models. Authors in [11] utilized a novel grey prediction model for the US, outperforming traditional ARIMA and LSTM benchmarks with conformable fractional accumulation. Similarly, authors in [12] applied a hybrid mathematical and ML approach to Canadian transportation data, forecasting a 50% rise in emissions by 2048, though their reliance on aggregated statistics potentially overlooked regional nuances. Extending this to the Asian region, authors in [13] optimized Neural Networks using Particle Swarm Optimization (PSO) to predict long-term trends. While these macro-level models demonstrate high accuracy, they often depend on forecasted input variables and lack the granularity required for component-level verification.

In complex industrial environments, dynamic and deep learning architectures have proven effective for real-time monitoring. Authors in [14] developed an adaptive regression framework for refinery biofeedstock co-processing, using change point detection to handle varying operating modes. Authors in [15] advanced this further by proposing a hybrid CNN-LSTM model within a "Metaverse" industrial waste control platform, achieving an R^2 of 0.995. However, while these distinct approaches highlight the capability of ML to handle nonlinear feature interactions, they often face challenges regarding generalizability and the scalability of conceptual environments, like the Metaverse, to broader regulatory contexts.

Specific to light-duty vehicles, research has benchmarked various algorithms to identify optimal predictors. Authors in [16] analyzed Canadian vehicle data, concluding that while engine size is a dominant factor, a simple Linear Regression model was sufficient to achieve an R^2 exceeding 99.7%.

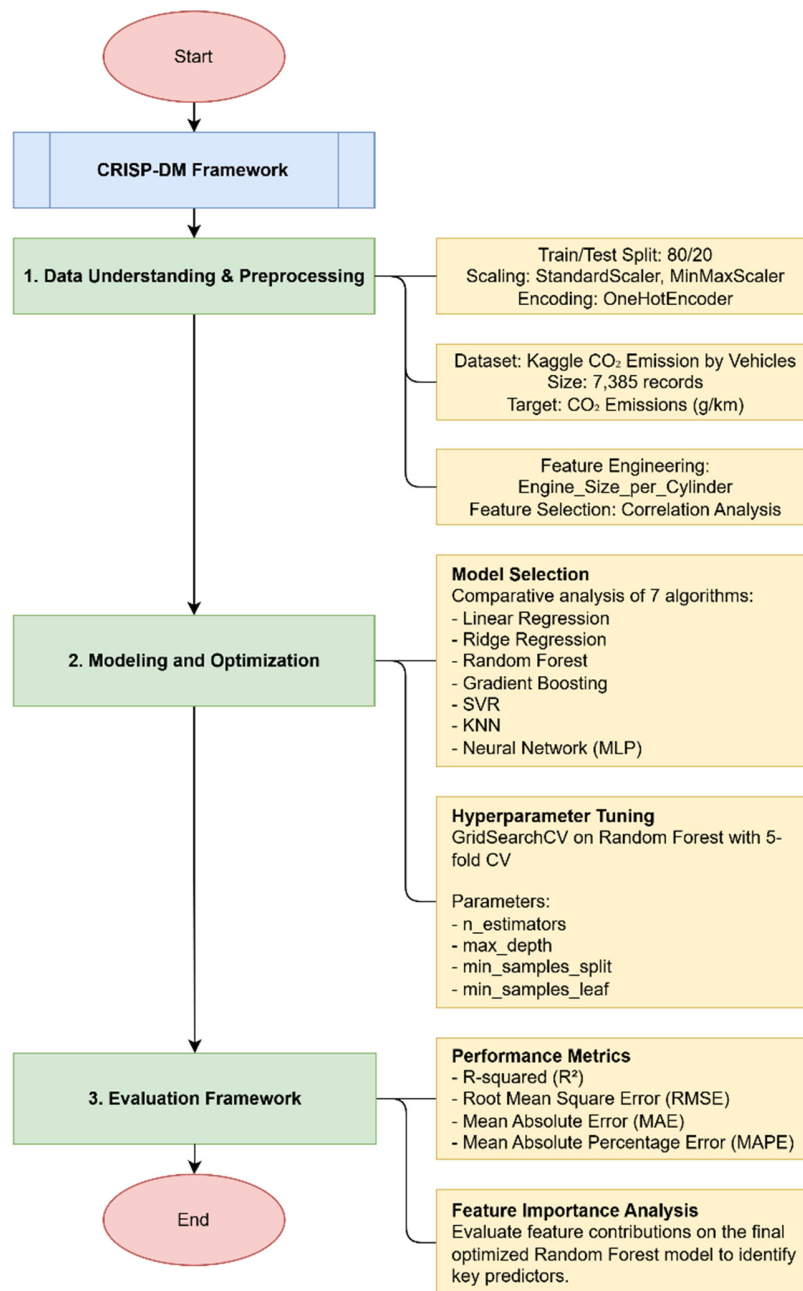


Fig. 1. Flowchart of the proposed framework.

The most recent literature from 2024 and 2025 has shifted focus from pure accuracy toward model interpretability and explainability. A 2025 study introduced an XAI framework for vehicle emissions, emphasizing that integrating interpretability layers (such as SHAP values) is crucial for regulatory adoption [17, 18]. Despite these advancements, a recurring limitation across the reviewed literature is the reliance on static datasets [16]. Most models do not yet fully account for real-world dynamic variables, such as road conditions and driving behavior, marking a critical gap that future research must

address to enhance the external validity of emission predictions.

III. PROPOSED METHOD

The methodology of this study adheres to the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, which ensures a structured and repeatable workflow encompassing data understanding, preparation, modeling, and evaluation. Figure 1 illustrates the overall research framework of the study.

A. Dataset and Preprocessing

A publicly available dataset was utilized, sourced from the Government of Canada's open data portal, titled 'CO₂ Emission by Vehicles' [19]. This dataset comprises 7,385 records of light-duty vehicles. Crucially, the dataset includes a measured target variable labeled 'CO₂ Emissions(g/km)' [20]. This variable represents the tailpipe carbon dioxide emissions for each vehicle, calculated based on the combined fuel consumption (a weighted average of 55% city and 45% highway driving). It serves as the ground truth for predictive modeling. The dataset also details independent features such as engine size (L), number of cylinders, transmission type, and fuel type. The correlation matrix between these numeric features is presented in Figure 2.

The target variable for prediction is CO₂ emissions, measured in g/km. The correlation between the fuel consumption and CO₂ emissions is displayed in Figure 3.

To enhance the model's predictive capability, a new feature, Engine_Size_per_Cylinder, was engineered by dividing the engine size by the number of cylinders to effectively capture engine efficiency. For feature selection, the absolute correlation of each input feature with the target variable was analyzed, and the six most correlated features were retained for modeling [21]. No features correlated above the 0.95 threshold, indicating that multicollinearity was not a significant concern. The correlation between engine size per cylinder and CO₂ emissions is portrayed in Figure 4.

The dataset was split into a training set of 5,907 records and a testing set of 1,477 records. For efficiency and optimal model performance, a conditional preprocessing pipeline was constructed. Numerical features for standard regression and distance-based models (Linear, Ridge, SVR, KNN) were scaled using StandardScaler (Z-score normalization). In contrast, specifically for the neural network model, the inputs were transformed using MinMaxScaler to the [0, 1] range to facilitate faster convergence and prevent vanishing gradient issues.

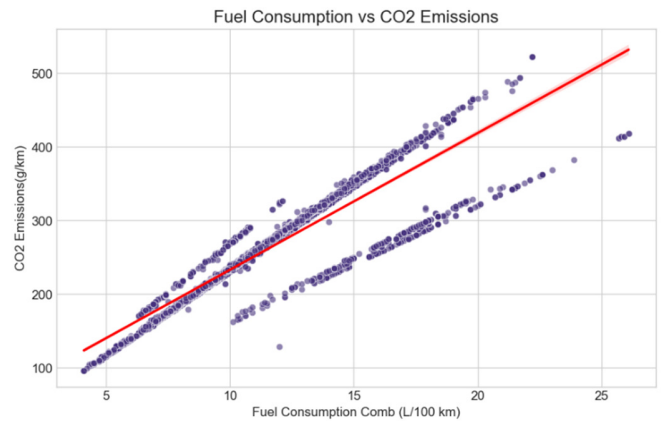


Fig. 3. Scatter plot between fuel consumption and CO₂ emissions.

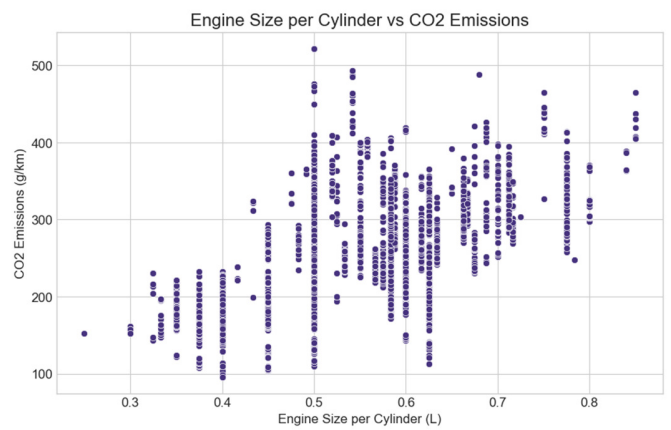


Fig. 4. Scatter plot between engine size per cylinder and CO₂ emissions.

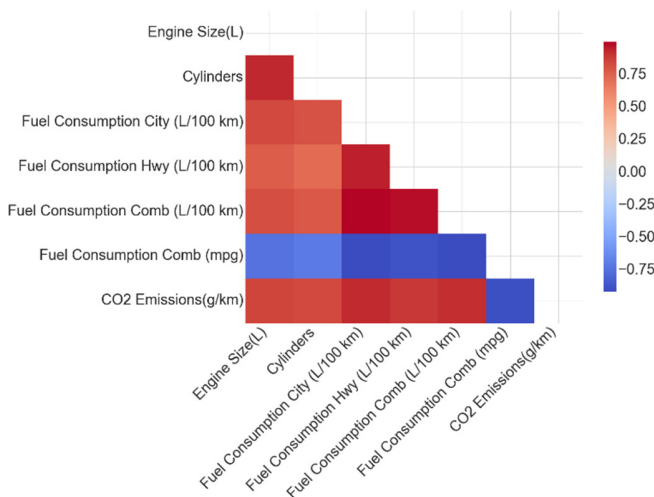


Fig. 2. Correlation matrix between numeric features.

B. Modeling and Optimization

A comparative analysis was conducted using seven supervised regression algorithms to identify the most effective model for predicting CO₂ emissions. The evaluated models included Linear Regression, Ridge Regression, Random Forest, Gradient Boosting, SVR, KNN, and a neural network (Multilayer Perceptron). All models were trained on the same preprocessed training data and evaluated on the unseen test set to ensure a fair and consistent comparison. To mitigate the effects of randomness and ensure reproducibility, a fixed random seed was applied to all algorithms and data splits during the experimental runs.

Based on the initial comparative results, the Random Forest model demonstrated the highest performance and was selected for further optimization. Hyperparameter tuning was performed using GridSearchCV with 5-fold cross-validation to find the optimal combination of model parameters [22]. The search grid focused on key hyperparameters, including n_estimators, max_depth, min_samples_split, and min_samples_leaf.

C. Evaluation Framework

The predictive performance of each model was quantitatively assessed using four standard regression metrics as defined and benchmarked in [23]. To measure the proportion of variance in CO₂ emissions explained by the model, the R²

metric was used, where a value closer to 1 signifies a better model fit:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (1)$$

To evaluate the model's error in the original units (g/km), the RMSE was calculated. RMSE represents the square root of the average of the squared differences between the actual (y_i) and predicted (\hat{y}_i) values and is particularly sensitive to significant errors:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

In addition, the Mean Absolute Error (MAE) was used to measure the average absolute differences between the predicted and observed values, providing a linear score where all individual errors are weighted equally:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3)$$

Finally, the Mean Absolute Percentage Error (MAPE) was employed to interpret the error in relative terms, which expresses the average absolute prediction error as a percentage of the actual values:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|A_i - F_i|}{A_i} \times 100\% \quad (4)$$

A feature importance analysis was conducted on the final optimized Random Forest model to enhance model interpretability. This technique evaluates the contribution of each input feature to the prediction accuracy by measuring its impact on the model's performance. This process identifies the most significant factors influencing vehicle CO₂ emissions, providing valuable insights for policy and manufacturing decisions.

IV. RESULTS AND DISCUSSION

A. Exploratory Data Analysis Findings

Initial exploratory data analysis revealed strong relationships between several vehicle specifications and the target variable, CO₂ emissions. The correlation analysis confirmed that Fuel Consumption Comb (L/100 km) had the highest positive correlation ($r = 0.92$) with CO₂ emissions, closely followed by Fuel Consumption City ($r = 0.92$). Other significant predictors included the Engine Size ($r = 0.85$) and the Number of Cylinders ($r = 0.83$). These findings underscore the primary role of fuel consumption and engine characteristics in determining a vehicle's emission levels.

B. Model Performance Comparison

The initial evaluation of the seven ML algorithms provided a clear performance hierarchy. As summarized in Table I, the ensemble methods—Random Forest and Gradient Boosting—along with the Neural Network, demonstrated the highest predictive accuracy. The Random Forest model emerged as the top performer with the highest initial R^2 and the lowest error metrics. In contrast, the SVR model yielded the lowest accuracy among all tested algorithms.

TABLE I. PERFORMANCE COMPARISON OF UNTUNED ML MODELS IN THE TEST DATA

Model	R^2	MAE (g/km)	RMSE (g/km)	MAPE (%)
Random Forest	1.0000	1.90	2.51	0.78
Gradient Boosting	0.9970	2.45	3.56	0.99
Neural Network	0.9950	2.54	4.04	1.05
KNN	0.9920	2.78	4.54	1.14
Linear Regression	0.9950	2.15	3.45	0.87
Ridge Regression	0.9950	2.18	3.46	0.88
SVR	0.9840	3.81	6.39	1.57

C. Optimized Model Performance

The Random Forest model was selected for hyperparameter tuning to further enhance its performance. The optimal parameters identified through GridSearchCV are presented in Table II.

TABLE II. OPTIMAL HYPERPARAMETERS FOR RANDOM FOREST REGRESSOR

Hyperparameter	Optimal value
n_estimators	200
max_depth	20
min_samples_split	2
min_samples_leaf	1

After applying these optimal parameters, the model's performance was re-evaluated on the test set. Table III shows that the tuned model achieved an R^2 of 0.9982 and an RMSE of 2.49 g/km, indicating an exceptionally accurate and robust predictive capability. This slight improvement over the already high-performing baseline model confirms the value of systematic optimization.

TABLE III. RANDOM FOREST PERFORMANCE BEFORE AND AFTER HYPERPARAMETER TUNING

Metric	Dataset	Before tuning	After tuning
R^2	Test	1.0000	0.9982
R^2	Train	1.0000	0.9994
RMSE (g/km)	Test	2.5100	2.4898
RMSE (g/km)	Train	1.3900	1.3803

Although the RMSE improved (decreased) from 2.51 g/km to 2.49 g/km after tuning, the R^2 value experienced a marginal decrease. This behavior is attributed to the regularization effect of the optimized hyperparameters, specifically the restriction of tree depth ($\text{max_depth} = 20$). The unconstrained 'default' Random Forest likely captured noise or specific outliers to maximize the variance explained (R^2), potentially leading to slight overfitting. In contrast, the tuned model prioritized the minimization of RMSE as the objective function. In the context of physical emission inventories, a reduction in the average RMSE magnitude is prioritized over a negligible drop in the correlation coefficient, as it reflects a more precise estimation of the actual pollutant mass. The benchmark comparison with recent studies on the same Canadian dataset is presented in Table IV.

TABLE IV. BENCHMARK COMPARISON WITH RECENT STUDIES ON THE SAME CANADIAN DATASET

Study	Methodology	R^2	RMSE (g/km)
Proposed method	Optimized Random Forest	0.9982	2.49
[18]	Gradient Boosting	0.9973	3.36
[16]	Linear Regression	0.997	N/A
[24]	Deep Learning (CarbonMLP)	0.9938	N/A

Table IV presents a direct quantitative comparison between the proposed optimized Random Forest model and recent state-of-the-art studies utilizing the same Canadian vehicle dataset. The results demonstrate that the proposed method establishes a new performance benchmark, achieving an R^2 of 0.9982 and an RMSE of 2.49 g/km, which surpasses the Gradient Boosting approach with $R^2 = 0.9973$ [18], the Linear Regression baseline with $R^2 = 0.9970$ [16], and the Deep Learning 'CarbonMLP' architecture with $R^2 = 0.9938$ [24]. This empirical superiority indicates that while deep learning and boosting techniques are powerful, the hyperparameter-tuned Random Forest algorithm is more effective at modeling the specific nonlinear interactions within this tabular domain, offering a more precise tool for regulatory verification with lower computational complexity.

D. Feature Importance

The feature importance analysis of the final optimized Random Forest model identified the most influential factors in predicting CO₂ emissions. The results confirmed that Fuel Consumption Comb (L/100 km) was the most important predictor. Other highly ranked features included Fuel Consumption Comb (mpg) and Fuel Type_Ethanol, corresponding to vehicles using E85 fuel.

E. Discussion

This study successfully developed and validated an ML model with exceptionally high accuracy for predicting CO₂ emissions from Canadian light-duty vehicles. The key finding is that an optimized Random Forest model achieved an R^2 of 0.9982, establishing a robust performance benchmark. The superior performance of the Random Forest model can be attributed to its ensemble nature, which combines multiple decision trees to capture complex, nonlinear interactions between features like engine size and fuel consumption that simpler models might miss.

To address concerns regarding potential overfitting and ensure model robustness, the 5-fold cross-validation procedure was analyzed for stability. The performance metrics remained consistent across the five random splits, indicating that the model's predictive power is not an artifact of a specific data partition. However, this validation is limited to the Canadian dataset. To fully establish the model's transferability, future validation phases must involve 'out-of-sample' testing on independent datasets from different geographic regions (e.g., U.S. or European fleet data) to confirm whether the optimized algorithm has similar performance across varying regulatory environments.

The findings are consistent with the broader literature, which consistently identifies ML as a powerful tool for emission prediction. For instance, the achieved accuracy

slightly exceeds the R^2 of 99.7% reported in [16], where the same dataset was used and a simpler Linear Regression model was found to be sufficient. This suggests that while linear relationships dominate this dataset, ensemble models such as Random Forest can capture residual nonlinearities to achieve state-of-the-art performance. In contrast to the national-level forecasting models developed in [13], the model provides granular, vehicle-specific predictions, making it more suitable for regulatory verification and component-level analysis.

For Canadian policymakers, this high-accuracy model is a powerful tool to verify manufacturer-reported data and simulate the effects of new emission standards. However, the integration of such predictive systems into national regulatory frameworks extends beyond technical accuracy; it raises critical questions regarding ethical AI and algorithmic accountability. To maintain public trust, the deployment of these models must prioritize transparency, ensuring that emission predictions are explainable and not merely 'black box' outputs. Policymakers must also rigorously audit for potential data biases, where a model might inadvertently penalize specific vehicle classes or regional driving patterns due to unbalanced training data. Therefore, these ML tools function as a preliminary 'digital audit' layer to flag anomalies for physical re-verification, rather than replacing traditional certification entirely. This hybrid approach ensures a fair, traceable compliance ecosystem that leverages AI efficiency while upholding rigorous environmental standards.

V. CONCLUSION

This study systematically evaluated seven Machine Learning (ML) models to predict the CO₂ emissions from Canadian light-duty vehicles, identifying the optimized Random Forest model as the superior predictor with an R^2 of 0.9982 and a Root Mean Square Error (RMSE) of 2.49 g/km. The novelty of this work lies in establishing a rigorous comparative benchmark that empirically demonstrates the necessity of ensemble learning methods over traditional linear models to accurately capture the nonlinear dependencies between engine specifications and fuel efficiency. These findings provide a robust, data-driven tool for regulators to verify manufacturer claims and audit compliance with greater precision than aggregate factor methods. While the current study is limited to static catalog data, this validated framework lays the foundation for future research to integrate dynamic real-world telemetry, thereby bridging the gap between laboratory certification benchmarks and on-road emission reality.

REFERENCES

- [1] M. Honegger, A. Michaelowa, and J. Roy, "Potential Implications of Carbon Dioxide Removal for the Sustainable Development Goals," *Climate Policy*, vol. 21, no. 5, pp. 678–698, May 2021, <https://doi.org/10.1080/14693062.2020.1843388>.
- [2] S. Hussain, E. Hussain, P. Saxena, A. Sharma, P. Thathola, and S. Sonwani, "Navigating the Impact of Climate Change in India: A Perspective on Climate Action (SDG13) and Sustainable Cities and Communities (SDG11)," *Frontiers in Sustainable Cities*, vol. 5, Jan. 2024, Art. no. 1308684, <https://doi.org/10.3389/frsc.2023.1308684>.
- [3] E. Vyshka, S. Sefa, F. Basholli, and D. Lumi, "Analysis of the Pollutant Emissions from Vehicles and their Impact on the Environment and Public Health," *Engineering, Technology & Applied Science Research*,

- vol. 15, no. 5, pp. 27660–27671, Oct. 2025, <https://doi.org/10.48084/etasr.12747>.
- [4] N. Kumarage, K. Hewage, S. Wanniarachchi, and R. Sadiq, "Canada's Transition to Light-duty Zero Emission Vehicles (ZEV): Opportunities, Challenges, and Policy Directions," *Transportation Research Part A: Policy and Practice*, vol. 199, Sept. 2025, Art. no. 104611, <https://doi.org/10.1016/j.tra.2025.104611>.
- [5] "CO₂ Emissions from Cars: Facts and Figures (Infographics)," *European Parliament*, Mar. 2019. <https://www.europarl.europa.eu/topics/en/article/20190313STO31218/co2-emissions-from-cars-facts-and-figures-infographics>.
- [6] Z. Long, J. Aksen, V. H. Gauer, and T. Niet, "Why Do Consumers Reject Zero-emissions Vehicles? Applying a Comprehensive Framework of Resistance to Canadian Car Buyers," *Energy Research & Social Science*, vol. 127, Sept. 2025, Art. no. 104297, <https://doi.org/10.1016/j.erss.2025.104297>.
- [7] V. Geetha Mohan, A.-F. Mubarak Ali, M. Ariff Ameen, B. Lathakumary Vijayan, A. Aminuddin, and W. Widayani, "Predictive Models Using Supervised Neural Network for Pollutant Removal Efficiency in Petrochemical Wastewater Treatment," in *2022 5th International Conference on Information and Communications Technology*, Yogyakarta, Indonesia, Aug. 2022, pp. 116–121, <https://doi.org/10.1109/ICOIACT55506.2022.9971929>.
- [8] A. Soltani, "Advanced Graph-based Machine Learning Reveals Cross-sector Drivers of Decarbonization in the United States and China," *Applied Energy*, vol. 397, Nov. 2025, Art. no. 126368, <https://doi.org/10.1016/j.apenergy.2025.126368>.
- [9] S. A. Khadem, F. Bensebaa, and N. Pelletier, "Optimized Feed-Forward Neural Networks to Address CO₂-Equivalent Emissions Data Gaps – Application to Emissions Prediction for Unit Processes of Fuel Life Cycle Inventories for Canadian Provinces," *Journal of Cleaner Production*, vol. 332, Jan. 2022, Art. no. 130053, <https://doi.org/10.1016/j.jclepro.2021.130053>.
- [10] "COPERT Methodology," *COPERT*, June 2025. <https://copert.emisia.com/copert/methodology/>.
- [11] S. Ding, X. Shen, H. Zhang, Z. Cai, and Y. Wang, "An Innovative Data-Feature-driven Approach for CO₂ Emission Predictive Analytics: A Perspective from Seasonality and Nonlinearity Characteristics," *Computers & Industrial Engineering*, vol. 192, June 2024, Art. no. 110195, <https://doi.org/10.1016/j.cie.2024.110195>.
- [12] M. Emami Javanmard, Y. Tang, Z. Wang, and P. Tontiwachwuthikul, "Forecast Energy Demand, CO₂ Emissions and Energy Resource Impacts for the Transportation Sector," *Applied Energy*, vol. 338, May 2023, Art. no. 120830, <https://doi.org/10.1016/j.apenergy.2023.120830>.
- [13] E. Abbasian Hamedani and S. Talebi, "Modeling and Long-Term Forecasting of CO₂ Emissions in Asia: An Optimized Artificial Neural Network Approach with Consideration of Renewable Energy Scenarios," *Energy Conversion and Management: X*, vol. 26, Apr. 2025, Art. no. 101030, <https://doi.org/10.1016/j.ecmx.2025.101030>.
- [14] L. Cao *et al.*, "Data-Driven Dynamic Modeling of Renewable CO₂ Emissions in Multimode Industrial Co-Processing Processes," *Control Engineering Practice*, vol. 164, Nov. 2025, Art. no. 106424, <https://doi.org/10.1016/j.conengprac.2025.106424>.
- [15] Y. Lin *et al.*, "A Sustainable Industrial Waste Control with AI for Predicting CO₂ for Climate Change Monitoring," *Journal of Environmental Management*, vol. 383, May 2025, Art. no. 125338, <https://doi.org/10.1016/j.jenvman.2025.125338>.
- [16] S. Boubaker, S. Al-Dahidi, and F. S. Alsubaei, "Modeling of CO₂ Emission for Light-Duty Vehicles: Insights from Machine Learning in a Logistics and Transportation Framework," *Computer Modeling in Engineering & Sciences*, vol. 143, no. 3, pp. 3583–3614, 2025, <https://doi.org/10.32604/cmescs.2025.063957>.
- [17] D. Yuan, L. Tang, X. Yang, F. Xu, and K. Liu, "Explainable Machine Learning Prediction of Vehicle CO₂ Emissions for Sustainable Energy and Transport," *Energies*, vol. 18, no. 20, Oct. 2025, Art. no. 5408, <https://doi.org/10.3390/en18205408>.
- [18] X. Guo, R. Kou, and X. He, "Towards Carbon Neutrality: Machine Learning Analysis of Vehicle Emissions in Canada," *Sustainability*, vol. 16, no. 23, Nov. 2024, Art. no. 10526, <https://doi.org/10.3390/su162310526>.
- [19] Government of Canada, "Fuel Consumption Ratings." Natural Resources Canada, Oct. 2013, [Online]. Available: <https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64>.
- [20] I. Oluwafemi, "CO₂ Emission of Vehicles in Canada." Kaggle, [online]. Available: <https://www.kaggle.com/datasets/isaacfemiogunniyi/co2-emission-of-vehicles-in-canada>.
- [21] R. S. Selina, M. Rahardi, A. Aminuddin, F. F. Abdulloh, H. Badi, and B. P. Asaddulloh, "Optimizing Diabetes Diagnosis Using Machine Learning with SMOTE and Feature Selection," in *2025 International Conference on Computer Sciences, Engineering, and Technology Innovation*, Jakarta, Indonesia, Jan. 2025, pp. 647–652, <https://doi.org/10.1109/ICoCSETI63724.2025.11020043>.
- [22] H. M. Khasanah, A. Aminuddin, F. F. Abdulloh, M. Rahardi, H. Hairani, and B. Pramudya, "Optimizing Mushroom Classification Through Machine Learning and Hyperparameter Tuning," *Engineering and Applied Science Research*, vol. 51, 2024, Art. no. 651660, <https://doi.org/10.14456/EASR.2024.61>.
- [23] D. Chicco, M. J. Warrens, and G. Jurman, "The Coefficient of Determination R-squared is More Informative Than SMAPE, MAE, MAPE, MSE and RMSE in Regression Analysis Evaluation," *PeerJ Computer Science*, vol. 7, July 2021, Art. no. e623, <https://doi.org/10.7717/peerj-cs.623>.
- [24] G. M. I. Alam *et al.*, "Deep Learning Model Based Prediction of Vehicle CO₂ Emissions with Explainable AI Integration for Sustainable Environment," *Scientific Reports*, vol. 15, no. 1, Jan. 2025, Art. no. 3655, <https://doi.org/10.1038/s41598-025-87233-y>.