Development of Fuel Consumption and CO_2 Emission Prediction Models Based on Driving Behavior Analysis Using GA-Optimized Curve Fitting

Sami Shaffiee Haghshenas, Sina Shaffiee Haghshenas, Giuseppe Guido, *Member*, *IEEE*, Vittorio Astarita, Maria Smyriliou

***Abstract*— Fuel consumption and CO_2 emissions in road transport are strongly affected by how a vehicle is operated under real driving conditions. This study develops prediction models for fuel consumption and CO_2 emissions using measurable vehicle-operation and driving-pattern indicators collected through On-Board Diagnostics (OBD) and smartphone-based tracking. The input variables include fuel flow rate per hour (FFRH), engine revolutions per minute (ERPM), speed (S), acceleration (A), and road grade (G). These variables are used as indirect indicators of driving patterns and operating conditions, rather than direct measurements of driver actions. A genetic algorithm (GA) is applied to optimize the coefficients of polynomial curve-fitting models and to obtain explicit mathematical equations for the two outputs. A dataset of 370 synchronized observations is used for model development. Several polynomial degrees and GA settings are examined to identify the most suitable model structure. The best fuel consumption model is obtained using a second-degree polynomial, while the CO_2 emission model requires a third-degree polynomial. The results show that GA-optimized curve fitting can provide interpretable prediction equations with acceptable accuracy. The proposed approach may be useful for eco-driving analysis, traffic management, and preliminary assessment of emission-reduction strategies.**

I. INTRODUCTION

Road transport is still a major contributor to fuel demand and CO_2 emissions, particularly in urban networks where frequent acceleration, deceleration, and stop-and-go movements occur [1], [2], [3]. For this reason, estimating fuel consumption and emissions at the vehicle level is important for traffic management, eco-driving assessment, and the early evaluation of emission-reduction strategies. Unlike aggregated approaches that mainly describe fleet-level trends, vehicle-level models can use operational data recorded during real trips and can therefore provide a closer representation of actual driving conditions.

Sami. S. H. Department of Civil Engineering, University of Calabria, Via Bucci, 87036 Rende, Italy (Corresponding author; e-mail: Sami.Shaffieehaghshenas@unical.it).

Sina. S. H. Department of Civil Engineering, University of Calabria, Via Bucci, 87036 Rende, Italy (e-mail: Sina. Shaffieehaghshenas@unical.it).

G. G. Department of Civil Engineering, University of Calabria, Via Bucci, 87036 Rende, Italy (e-mail: Giuseppe.guido@unical.it).

V. A. Department of Civil Engineering, University of Calabria, Via Bucci, 87036 Rende, Italy (e-mail: Vittorio.astarita@unical.it).

M. S. Department of Chemistry and Chemical Technologies, University of Calabria, 87036 Arcavacata di Rende, Italy (e-mail: Smyrilioumaria@gmail.com).

In this study, the term driving pattern is used to describe measurable vehicle-operation indicators obtained from vehicle response during a trip. These indicators include fuel flow rate per hour, engine revolutions per minute, speed, acceleration, and road grade. They should not be interpreted as direct measurements of driver actions, such as pedal position, steering input, reaction time, or cognitive state. Instead, they are considered observable proxies that reflect how the vehicle is operated under real-world conditions. This distinction is important because OBD- and smartphone-based data acquisition systems mainly record vehicle dynamics and engine-related variables, rather than the driver's physical or cognitive actions [4], [5].

Several studies have examined fuel consumption, transport energy use, and CO_2 emissions from different viewpoints. Some works have focused on the effects of ride-sourcing services, vehicle ownership, travel demand, and related emissions [6]. Other studies have investigated the role of fuel-consumption regulations, fleet electrification, and policy scenarios in reducing greenhouse gas emissions from road transport [7], [8]. In the public transport sector, alternative technologies such as natural gas, battery-electric, and hydrogen buses have also been compared in terms of cost, energy use, and emission impacts [9]. These studies show that transport emissions are affected by a combination of vehicle technology, travel activity, policy measures, and operating conditions.

Alongside policy- and fleet-level studies, data-driven methods have increasingly been used to predict fuel consumption and emissions from vehicle and traffic variables. Machine-learning models can capture nonlinear relationships in operational data, but many of them work as black-box tools and do not provide simple mathematical relations between input variables and predicted outputs [4], [5]. This limits their use when the objective is not only prediction, but also interpretation of how vehicle-operation indicators influence fuel consumption and CO_2 emissions. In addition, in some modelling frameworks, the structure of the prediction function and its coefficients are not optimized in a systematic way, which may affect model stability and reproducibility.

Based on these gaps, this study develops GA-optimized polynomial curve-fitting models for predicting fuel consumption and CO_2 emissions using vehicle-operation and driving-pattern indicators. The purpose is to obtain explicit prediction equations whose coefficients are optimized through a genetic algorithm, while keeping the model structure interpretable. The proposed framework uses fuel flow rate per hour, engine RPM, speed, acceleration, and road grade as inputs, and develops separate models for fuel consumption and CO_2 emissions. By focusing on measurable vehicle-response variables rather than direct driver-action measurements, the study provides a clearer and more reproducible modelling framework for real-world vehicle-operation analysis.

The study is intended as an interpretable curve-fitting framework for the available real-driving dataset, rather than as a complete benchmark against all existing prediction methods.

II. Methodology

A. Curve Fitting Technique

Curve fitting was used in this study to obtain explicit equations between the input variables and the target outputs. Unlike black-box models, this approach gives a mathematical form that can be inspected and later used for interpretation. The input vector includes fuel flow rate per hour (FFRH), engine revolutions per minute (ERPM), speed (S), acceleration (A), and road grade (G). The target variable is either fuel consumption (FC) or CO_2 emission [10], [11].

A polynomial function was selected because the relation between vehicle-operation variables and the two outputs is not expected to be purely linear. Polynomial terms allow the model to include nonlinear effects and interactions between variables, while still keeping the final equation readable. The general form of the polynomial model can be written as Eq (1) [12], [13].

$$Y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + ... + a_n x^n \tag{1}$$

B. Genetic Algorithm

In transportation forecasting problems, many relationships between variables are nonlinear and complex in nature, and for this reason, machine learning and deep learning methods are widely used to extract hidden patterns from data. However, a significant part of the success of these models depends on how their structure and parameters are adjusted, and without a proper search process, it will be difficult to reach the optimal solution. Hence, the use of intelligent optimization methods that can efficiently explore the response space becomes particularly important [14], [15], [16], [17].

The Genetic Algorithm (GA) is a powerful evolutionary algorithm widely used in artificial intelligence to evaluate complex systems [18]. The genetic algorithm, introduced by John Holland and his colleagues in the 1960s, is based on stochastic search techniques and is inspired by biological evolution [19], [20]. Its applications span across various engineering fields [21], [22]. The GA process consists of three main operators: reproduction, crossover, and mutation. In the first step, an initial population is randomly generated from different solution representations, usually in the form of numeric strings. Then, using the fitness function, the quality and efficiency of each solution is assessed. In the reproduction stage, the most suitable solutions are selected as parents to create new generations through the crossover operator. The mutation operator helps to improve the fitness values compared to the previous generation by applying random changes to the new generation. This process continues iteratively until the desired level of optimization is achieved and finally, the process is stopped.

III. Driving Behavior Characteristics and Their Influence on Fuel Consumption and CO_2 Emissions

Drivers' behavior is one of the most important factors that directly affects fuel consumption and the amount of CO_2 emitted. Other factors such as speed, acceleration, and FFRH and ERMP have an impact on fuel consumption and greenhouse gas emissions. That is, as speed increases, fuel consumption and CO_2 emissions increase, but with more stable speed and acceleration patterns, fuel consumption decreases, which in turn reduces CO_2 emissions. With data obtained from vehicle sensors, driver behavior can be analyzed, and predictive models can be developed. Therefore, understanding and recognition the relationship between driver behavior and fuel consumption is a critical step towards developing accurate and reliable models for estimating fuel consumption and CO_2 emissions in real driving conditions [23], [24], [25].

IV. Results

In this study, a dataset with 370 data points is used for modeling of fuel consumed (FC) and CO_2. Since the dataset is limited in size and was obtained from a small group of vehicles, the present analysis is treated as a model-fitting study rather than a full generalization test. The reported error values describe how well each optimized polynomial form follows the available observations. They should not be read as a final estimate of model performance for all vehicle types, routes, or traffic conditions. A wider validation using independent trips and additional vehicles is therefore required before applying the equations beyond the current data range. The input data for modeling and determining an optimized function for predicting include Fuel_flow_rate/hour (FFRH), Engine_RPM (ERMP), Speed (S), Acceleration (A), and Grade (G). Also, the amount of fuel consumed (FC) and CO_2 are also considered output data respectively. The adopted dataset pertains to measurements of certain parameters obtained from a sample of private vehicles (3 cars and 1 commercial vehicle) through On-Board Diagnostics (OBD) interface and smartphones onboard. It should be mentioned that the data acquisition system is based on the sharing of vehicle track data acquired by mobile technology and OBD interface and sent by users to a central server which processes and organizes them according to the different points of interest. The system is therefore composed of two parts: 1) a client, composed of a mobile device which also collects the data transmitted via Bluetooth from the OBD interface, and 2) a central web server, based on the Linux-MySQL-PHP combination, which processes them and returns them synchronized information [26].

A. Modeling and Results for Fuel Consumption

In this modeling, to achieve the best solution, the control parameters of the genetic algorithm must be selected first. For this purpose, although there are no specific relationships, it is possible to determine a range of each of the control parameters using past studies and experts' opinions and determine their best value by trial and error. For this purpose, a range of population values is considered for the genetic algorithm, including 100, 200, 300, 400 and 500 as well as the number of iterations is also considered, including 100, 300, 500, 1000, and 1500 iterations. Therefore, 25 models are established, which are considered for polynomials with degrees 1 to 4. As a result, a total of 100 models were made, and their results were analyzed. After evaluating the tested configurations, the selected configuration with the lowest fitting error among the tested polynomial-GA settings was obtained using a second-degree polynomial, a population size of 300, and 1500 iterations, as presented in Eq. (2). Figures 1 to 3 show the accuracy of the model, error rate, and curve fitting, as well as the performance of the model in each iteration.

$$
\begin{aligned}
FC = & -0.99-(FFRH)+0.01(FFRH)^2+0.96(ERMP)\ldots \\
& -(ERMP)^2+0.34(S)-0.01(S)^2-0.01(A)^2\ldots \\
& +0.82(G)-0.42(G)^2
\end{aligned}
\tag{2}
$$

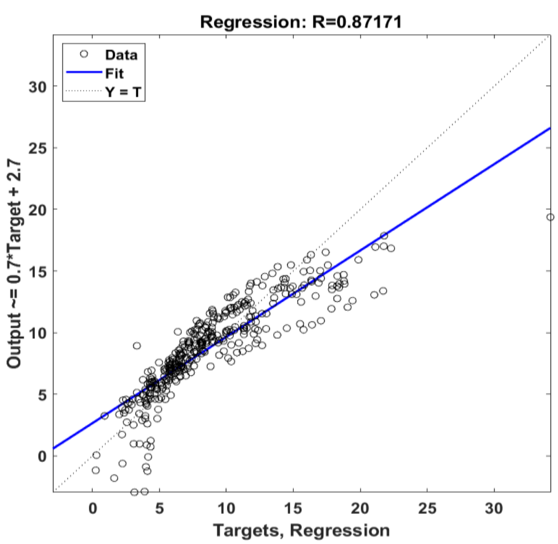

Figure 1. Regression result of the best model for fuel consumption.

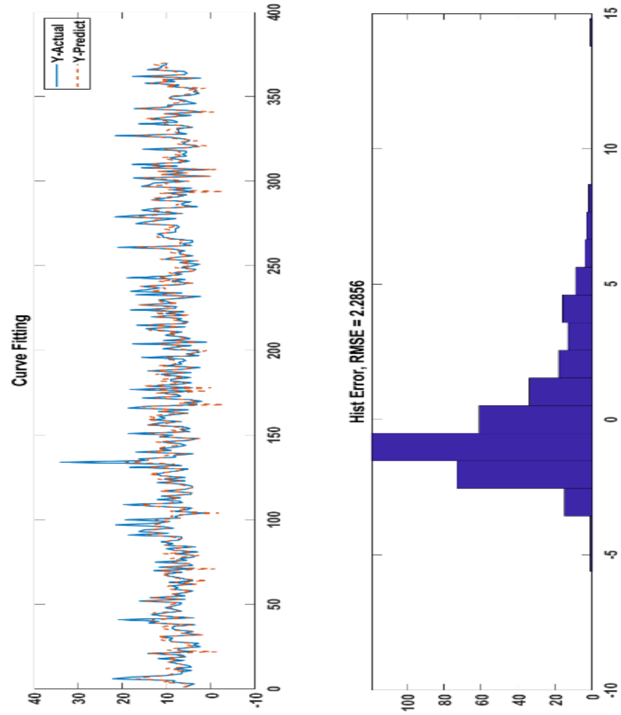

Figure 2. Curve fitting and RMSE results of the best model for fuel consumption.

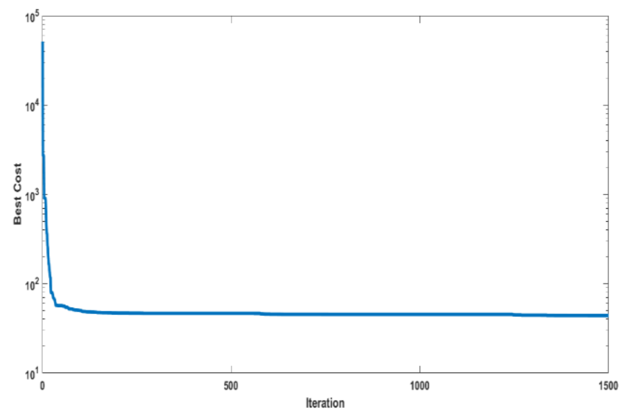

Figure 3. Performance process of the best model in each iteration.

B. Modeling and Results for CO_2 Emission

The first step is to find the best control parameters of the genetic algorithm to optimize the curve fitting technique, which can provide us with an optimized formula with the best accuracy for predicting the rate of CO_2 emissions. Therefore, considering a range similar to the modeling of the previous section for the number of populations and replication, 25 initial models were created. Also, in this modeling, polynomials with degrees up to 4 were considered, and a total of 100 models were made. After comparing the examined configurations, the configuration that gave the lowest fitting error within the examined search range was obtained using a third-degree polynomial, a population size of 400, and 1500 iterations, as presented in Eq. (3). Figures 4 to 6 show the accuracy of the model, the error rate, the curve fit, the error of the model, and the performance of the modeling process in each iteration, respectively.

$$\begin{aligned}
CO_2 &= 1+(FFRH)+0.43(FFRH)^2-(FFRH)^3 \ldots \\
&+(ERMP)+(ERMP)^2+(ERMP)^3-0.0001(S)\ldots \\
&-(S)^2+(S)^3+(A)+0.1(A)^2\ldots \\
&-0.002(G)+(G)^2-(G)^3
\end{aligned} \tag{3}$$

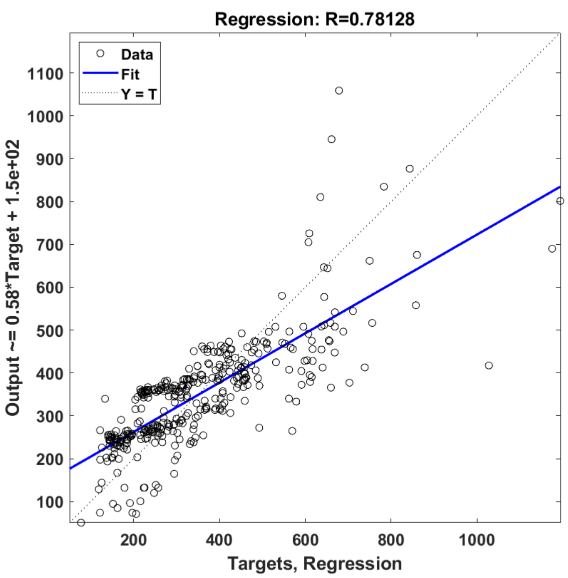

Figure 4. Regression result of the best model for CO_2 emission.

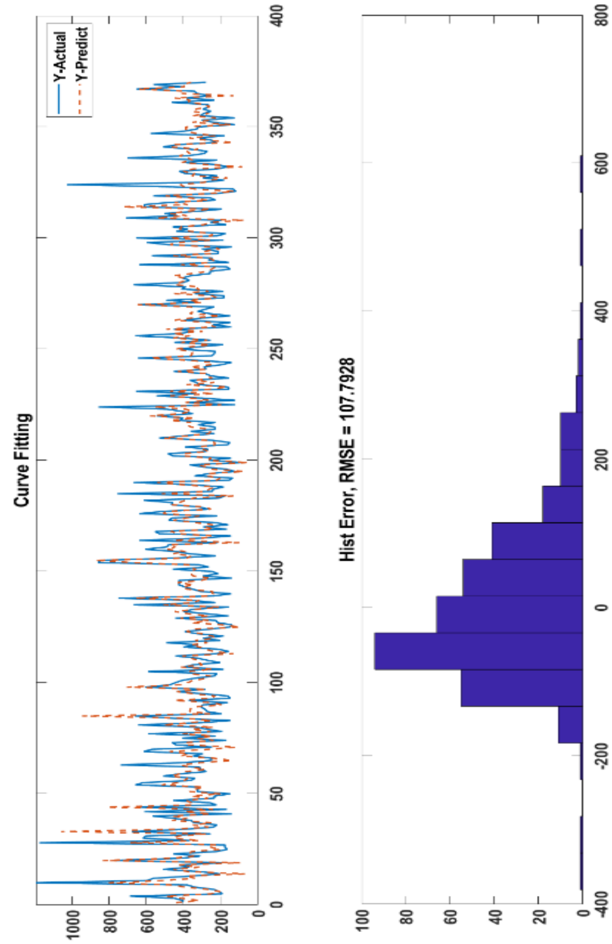

Figure 5. Curve fitting and RMSE results of the best model for CO_2 emission.

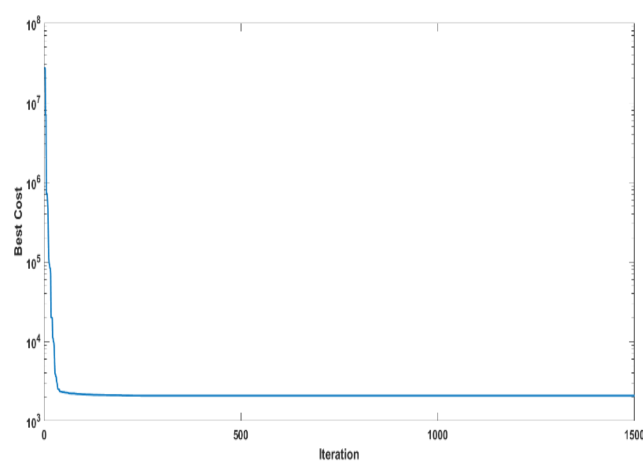

Figure 6. Performance process of the best model in each iteration for CO_2 emission.

C. Discussion

In this study, an attempt was made to provide an interpretable model for predicting fuel consumption and CO_2

emissions by relying on driving behavior analysis and applying optimized curve fitting by genetic algorithm. Unlike many studies that only use black box models, here the focus was on extracting an analytical relationship between measurable vehicle parameters and energy and emission outputs so that the results could be used in transportation management and driver training. Data were collected through OBD and mobile phone systems and under real driving conditions and included variables such as Fuel_flow_rate/hour (FFRH), Engine_RPM (ERMP), Speed (S), Acceleration (A), and Grade (G), therefore, the model is close to real conditions and is not simply the result of laboratory data. Overall, the results of this study showed that driver behavior plays an important role in fuel consumption and pollutant production, and even small changes in the vehicle movement pattern can make a significant difference in energy output. Overall, the main insight of this research is that before moving towards costly technologies, a practical step towards sustainable transportation can be taken by understanding and modifying driving behavior and providing accurate predictive models.

The results of the presented images and graphs show that the optimization process played a decisive role in the accuracy of the model. In the fuel consumption model, the best response was obtained from a second-degree polynomial with a population of 300 and 1500 iterations, while for CO_2 emissions, the third-degree model with a population of 400 and 1500 iterations showed a more appropriate performance. This difference indicates that the complexity of the pollutant emission phenomenon is greater than that of fuel consumption and that higher mathematical flexibility is required to describe it. The aim of this work is not to rank the proposed model against all possible prediction methods. More complex models, such as neural networks or ensemble learning methods, may produce lower errors when trained on larger datasets. However, they often provide limited access to the mathematical relation between inputs and outputs. The polynomial-GA approach was selected because it produces an explicit equation and keeps the modelling process easier to inspect. Therefore, the value of the proposed model is mainly related to transparency and compact formulation, while a wider benchmark against alternative methods remains necessary in future work.

The trend of error reduction in the iteration graphs also shows that the genetic algorithm has gradually searched the response space and converged to a stable point; that is, instead of getting stuck in a local minimum, it has been able to find the best combination of coefficients. Also, the shape of the curve fitting shows that the model behaves logically not only in the intermediate data but also in the boundary ranges and has not undergone severe fluctuations, which is very important for practical application. From an optimization perspective, the appropriate choice of population size and number of iterations created a balance between computational cost and accuracy; smaller populations were faster but less accurate, and larger populations were costly. Thus, the important conclusion of this section is that the genetic algorithm itself is not the only factor in success, but that tuning its parameters is a key part of the modeling process and can significantly change the quality of the prediction.

It is clear that as the speed increases, the engine speed increases, which causes fuel consumption and CO_2 production to increase. On the contrary, uniform movement and slow acceleration reduce energy consumption. For this reason, the presented model can be used practically in urban traffic management; for example, determining appropriate speeds, preventing frequent stops, and making the flow of vehicles smoother can reduce fleet fuel consumption. These results show that urban pollution does not only depend on the type of vehicle and driving style also has a significant contribution, so driver training or the use of intelligent driving systems can be effective alongside technical solutions. In addition, having an accurate prediction model allows for the estimation of the effect of projects such as smart traffic lights or route modifications on fuel consumption and CO_2 emissions before implementing them. Overall, it can be said that modifying driving patterns is one of the least expensive ways to improve energy efficiency in transportation. However, it should be noted that the limited number of data and the variety of vehicles studied can affect the generalizability of the results.

V. Conclusion

This study sought to paint a tangible picture of the relationship between driving style, fuel consumption and CO_2 emissions, rather than simply presenting a statistical relationship. Many studies rely on models that only predict the outcome but do not provide a clear explanation of why; therefore, a combination of curve fitting and genetic algorithms was used here to obtain a model that is both reasonably accurate and interpretable. Real data extracted from cars also made the model close to practical conditions and enabled its use in traffic management, driver training, and the evaluation of control policies. Overall, the findings show that modifying driving patterns can be a simple and effective solution to reduce energy consumption and pollution before costly technological investments. In future work, expanding the database with a greater variety of vehicles and utilizing real-time urban data could increase the accuracy and generalizability of the model. The findings should be interpreted within the limits of the available dataset. The equations were derived from a restricted number of observations and vehicles, and their use under other traffic, road, or vehicle conditions requires additional testing. Future studies should examine independent trips, include a more diverse vehicle sample, and compare the proposed formulation with other statistical and machine-learning models.

References

[1] S. S. Haghshenas, V. Astarita, S. S. Haghshenas, and G. Guido, "Artificial intelligence-powered driver behavior analysis for fuel consumption optimization: a pathway to greener roads," in *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, IEEE, 2024, pp. 116–122.

[2] S. Jafarzadeh Ghoushchi, S. Shaffiee Haghshenas, S. Vahabzadeh, S. Shaffiee Haghshenas, V. Astarita, and G. Guido, "Risk assessment of young driver behavior using an extended decision-making approach based on FMEA in uncertain

environments," *Neural Comput. Appl.*, pp. 1–38, 2025.

[3] S. S. Haghshenas, V. Astarita, S. S. Haghshenas, G. Guido, and A. Kouvelas, "A New Perspective on Artificial Intelligence Applications in Analyzing Driver Behavior: Advances, Challenges, and Opportunities," in *2025 11th International Conference on Control, Decision and Information Technologies (CoDIT)*, IEEE, 2025, pp. 1107–1112.

[4] S. S. Haghshenas, V. Astarita, S. S. Haghshenas, G. Guido, and S. J. Ghoushchi, "Review of applications of ML approaches in driver behavior analysis using qualitative and quantitative analysis," in *2023 9th International conference on control, decision and information technologies (CoDIT)*, IEEE, 2023, pp. 1207–1213.

[5] M. H. Mobini Seraji *et al.*, "A state-of-the-art review on machine learning techniques for driving behavior analysis: clustering and classification approaches," *Complex Intell. Syst.*, vol. 11, no. 9, p. 386, 2025.

[6] J. W. Ward, J. J. Michalek, I. L. Azevedo, C. Samaras, and P. Ferreira, "Effects of on-demand ridesourcing on vehicle ownership, fuel consumption, vehicle miles traveled, and emissions per capita in US States," *Transp. Res. Part C Emerg. Technol.*, vol. 108, pp. 289–301, 2019.

[7] F. Zhao, F. Liu, Z. Liu, and H. Hao, "The correlated impacts of fuel consumption improvements and vehicle electrification on vehicle greenhouse gas emissions in China," *J. Clean. Prod.*, vol. 207, pp. 702–716, 2019.

[8] M. Asim *et al.*, "Estimating the long-term effects of national and international sustainable transport policies on energy consumption and emissions of road transport sector of Pakistan," *Sustainability*, vol. 14, no. 9, p. 5732, 2022.

[9] P. Muñoz, E. A. Franceschini, D. Levitan, C. R. Rodriguez, T. Humana, and G. C. Perelmuter, "Comparative analysis of cost, emissions and fuel consumption of diesel, natural gas, electric and hydrogen urban buses," *Energy Convers. Manag.*, vol. 257, p. 115412, 2022.

[10] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *J. ACM*, vol. 17, no. 4, pp. 589–602, 1970.

[11] P. G. Guest and P. G. Guest, *Numerical methods of curve fitting*. Cambridge University Press, 2012.

[12] C. M. Bishop and C. M. Roach, "Fast curve fitting using neural networks," *Rev. Sci. Instrum.*, vol. 63, no. 10, pp. 4450–4456, 1992.

[13] T. Nakazawa, M. Ishizawa, K. A. Z. Higuchi, and N. B. A. Trivett, "Two curve fitting methods applied to CO2 flask data," *Environmetrics Off. J. Int. Environmetrics Soc.*, vol. 8, no. 3, pp. 197–218, 1997.

[14] N. Shamsmohammadi, S. Shaffiee Haghshenas, G. Guido, and V. Astarita, "Enhancing Operational Efficacy of Smart Parking Facilities through Intelligent License Plate Recognition," *J. Soft Comput. Civ. Eng.*, vol. 8, no. 4, pp. 94–115, 2024.

[15] V. Astarita, G. Guido, S. S. Haghshenas, S. S. Haghshenas, and G. Martino, "Risk reduction in transportation systems: A Bibliometric-Based Literature Review on Artificial Intelligence in Disaster Response," *Transp. Res. Procedia*, vol. 90, pp. 599–606, 2025.

[16] S. S. Haghshenas, G. Guido, S. S. Haghshenas, and V. Astarita, "Predicting the level of road crash severity: A comparative analysis of logit model and machine learning models," *Transp. Eng.*, vol. 20, p. 100323, 2025.

[17] S. S. Haghshenas, G. Guido, S. S. Haghshenas, V. Astarita, and V. Simic, "Artificial intelligence techniques in road safety for sustainable mobility: an investigation into applications, challenges, and opportunities," *Reliab. Decis. Sustain. Transp.*, pp. 203–223, 2026.

[18] A. Aryafar, R. Mikaeil, S. S. Haghshenas, and S. S. Haghshenas, "Application of metaheuristic algorithms to optimal clustering of sawing machine vibration," *Measurement*, vol. 124, pp. 20–31, 2018.

[19] J. H. Holland, *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[20] G. Guido, S. S. Haghshenas, S. S. Haghshenas, A. Vitale, V. Astarita, and A. S. Haghshenas, "Feasibility of stochastic models for evaluation of potential factors for safety: a case study in Southern Italy," *Sustainability*, vol. 12, no. 18, p. 7541, 2020.

[21] O. E. Canyurt and H. K. Ozturk, "Application of genetic algorithm (GA) technique on demand estimation of fossil fuels in Turkey," *Energy Policy*, vol. 36, no. 7, pp. 2562–2569, 2008.

[22] A. Gálvez and A. Iglesias, "A new iterative mutually coupled hybrid GA–PSO approach for curve fitting in manufacturing," *Appl. Soft Comput.*, vol. 13, no. 3, pp. 1491–1504, 2013.

[23] M. Karrouchi *et al.*, "Driving behavior assessment: a practical study and technique for detecting a driver's condition and driving style," *Transp. Eng.*, vol. 14, p. 100217, 2023.

[24] E. Bagheri, M. M. Tehrani, M. Azadi, and A. Moosavian, "Impact of driving characteristic parameters and vehicle type on fuel consumption and emissions performance over real driving cycles," *PLoS One*, vol. 20, no. 1, p. e0317098, 2025.

[25] S. S. Haghshenas, S. S. Haghshenas, V. Astarita, G. Guido, and M. Mohamed, "Driving toward efficiency: analysis of driver behavior and fuel consumption through machine learning," in *Reliable Decision-Making for Sustainable Transportation*, Elsevier, 2026, pp. 275–297.

[26] S. S. Haghshenas, S. S. Haghshenas, V. Astarita, and G. Guido, "Fuel Consumption and CO_2 Emissions Prediction in Road Transport Using a Hybrid Deep Learning Approach," *Transp. Eng.*, p. 100364, 2025.