Discover Sustainability

Research

Incorporating driving behavior into vehicle fuel consumption prediction: methodology development and testing

Huthaifa I. Ashqar[1,2] · Mahmoud Obaid[3] · Ahmed Jaber[4] · Rashed Ashqar[5] · Nour O. Khanfar[6] · Mohammed Elhenawy[7]

Received: 21 June 2024 / Accepted: 19 September 2024
Published online: 22 October 2024

Abstract

This study is a significant endeavor involving the development and testing of a comprehensive methodology to incorporate driving behavior into the analysis and prediction of vehicle fuel consumption. It underscores the crucial importance of understanding how different driving behavior affect fuel efficiency. The framework we present is a theoretical construct and a practical tool. It provides a robust, multi-step process for linking driving behavior to fuel consumption, leveraging both traditional statistical methods and advanced machine learning techniques to derive actionable insights. To test the framework, we used a naturalistic data that includes about 5408 different road users in a mixed traffic environment and urban settings in Germany. We applied a microscopic fuel consumption model to calibrate the framework and an unsupervised clustering algorithm to classify the behavior of the driver interacting with each other and with vulnerable road users. The framework includes developing Linear regression model as a baseline, which yields an R-squared of 0.511 and a Mean Squared Error (MSE) of 0.031, indicating moderate predictive accuracy. The final step includes choosing Random Forest as a better model, which achieves a higher R-squared of 0.956 and a lower MSE of 0.003. We also found that conservative and aggressive driving leads to significantly higher and more discrepancy in fuel consumption than normal driving behavior. These insights can promote more efficient driving practices, leading to significant fuel savings and environmental benefits.

Article highlights

- Driving behavior significantly impacts fuel consumption; smooth, controlled driving leads to greater fuel efficiency.
- Insights from this study can inform ADAS development, enabling real-time, fuel-efficient driving assistance systems.
- Promoting normal driving habits can reduce fuel use, aiding eco-friendly driving and sustainable transport solutions.

✉ Ahmed Jaber, ahjaber6@edu.bme.hu; Huthaifa I. Ashqar, huthaifa.ashqar@aaup.edu; Mahmoud Obaid, Mahmoud.Obaid@aaup.edu; Rashed Ashqar, rashedessam2@gmail.com; Nour O. Khanfar, n.khanfar1@student.aaup.edu; Mohammed Elhenawy, mohammed.elhenawy@qut.edu.au | [1]Civil Engineering Department, Arab American University, 13 Zababdeh, P.O Box 240, Jenin, Palestine. [2]Artificial Intelligence Program, Columbia University, 500 W 120th St, New York, NY 10027, USA. [3]Computer System Engineering Department, Arab American University, 13 Zababdeh, P.O Box 240, Jenin, Palestine. [4]Department of Transport Technology and Economics, Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Műegyetem rkp. 3, Budapest 1111, Hungary. [5]Finance Department, Al Zaytona University of Science & Technology, Salfit Street, Al-Laban Al-Sharqiya, Salfit, Palestine. [6]Natural, Engineering and Technology Sciences Department, Arab American University, 13 Zababde, P.O. Box 240, Jenin, Palestine. [7]Accident Research and Road Safety Queensland, Queensland University of Technology, 130 Victoria Park Rd, Kelvin Grove, Brisbane, QLD 4059, Australia.

Keywords Driver behavior · Fuel consumption · Mixed traffic · Naturalistic data · Volatility measures

1 Introduction

Vehicle fuel consumption prediction is an important factor that can help vehicle owners to reduce their fuel costs. The prediction of fuel consumption can be done through various methods. One of the most popular methods is based on driving behavior, which uses data from sensors in the vehicle to measure and analyze the driver's driving style. This data can then be used to predict the vehicle's fuel consumption. The driving behavior method of fuel consumption prediction is because the way drivers drive their vehicles affects fuel consumption. The data includes the vehicle's speed, acceleration, braking, and cornering levels. Based on this data, the system can determine how much fuel the vehicle will consume under different driving conditions. The driving behavior method of fuel consumption prediction has been tested in several studies and is accurate and reliable. For example, a study conducted by Vyas et al. [1] found that the driving behavior method was able to predict fuel consumption with an accuracy of 95%. This method is also useful for predicting the fuel economy of different types of vehicles, such as cars, SUVs, and trucks. Overall, the driving behavior method of fuel consumption prediction is a reliable and accurate way to predict fuel consumption. This method can be used to help vehicle owners reduce their fuel costs, as well as to improve the fuel economy of their vehicles.

Due to advancements in information and communication technology, it is now possible to gather vast amounts of driving data online to monitor the activity of vehicle fleets and ultimately increase fleet management performance [2]. Solving vehicle routing problems (VRP) is an essential component of fleet management systems [3]. This is done to assign vehicles and routes for accomplishing driving/delivery missions with a minimum number of vehicles and fuel costs. To optimize the VRP, predicting vehicle fuel consumption is necessary, taking into consideration the given route, period of day, driver, etc [4].

A model is used to predict fuel consumption, considering both driving behaviors and traffic conditions. However, certain variables that influence fuel consumption, such as vehicle weight and ambient conditions, cannot be measured or broadcasted from onboard systems. This means that fuel consumption must be modeled using only a limited number of keys and standard variables, including vehicle velocity and road slope [5]. In addition to vehicle velocity and road slope, vehicle acceleration plays a crucial role in determining fuel consumption. Variability in acceleration, particularly in stop-and-go traffic, has a significant impact on fuel efficiency. Fuel consumption models can be broadly classified into two types: (I) first-principle (physics-based) models that utilize a set of mathematical equations to simulate vehicle dynamics at each moment in time [6], and (ii) data-driven machine learning (ML) models that use a set of input/explanatory variables to create a mapping to an output space defined by target variable(s) [7].

Moreover, physics-based approaches can be accurate, but they often lack computing efficiency [8, 9]. Additionally, they require knowledge of various vehicle dynamics, powertrain parameters, and multidimensional maps, which are usually unavailable. To make VRP feasible, there is a need for a fast, macroscopic model that can accurately predict fuel consumption for an entire driving cycle. ML modeling approaches can be used to solve this problem, as they allow for automatic pattern learning from available data [10]. The most used models are random forest (RF) and artificial neural networks (NNs), which are universal approximators that can represent the nonlinear characteristics of a complex system through a nonlinear activation function [11].

In recent times, the utilization of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) in combination with deep learning (DL) techniques has led to a great improvement in the success rate of NNs for regression and classification tasks. A thorough analysis of different advanced ML models and DL techniques is provided, with the prediction accuracies of corresponding models being tested and compared in the context of a short-term electric microgrid load forecasting problem [12]. If a more intricate prediction job is to be carried out, it is necessary to consider probabilistic model output, which, instead of point predictions (i.e., related to expected values), requires the prediction of conditional probability distributions of dependent variables or related statistical indices (e.g., quantiles) [13].

Driving behavior significantly impacts vehicle fuel consumption, presenting an opportunity to reduce transportation energy costs and enhance the technology for behavior assessment in Advanced Driver Assistance Systems (ADAS). However, understanding and modeling driving behavior under dynamic and mixed traffic conditions is complex, making it challenging to quantitatively analyze the relationship between driving behavior and fuel consumption. Despite this complexity, accurately predicting fuel consumption based on driver behavior is essential for developing more ecological driving assistance systems and improving vehicle fuel economy. The research gap addressed in this study lies in the lack of integrated models that combine both driving behavior and environmental factors for real-time fuel consumption prediction. Our contribution is a comprehensive multi-step framework that links driving volatility measures with fuel consumption, validated using naturalistic driving data. This study aims to address these challenges by proposing a method that effectively identifies the relationship between driving behavior and fuel consumption at both macro and micro levels, enabling real-time fuel consumption calculations without the need for storing instantaneous data. This approach advances the understanding of driver behavior and facilitates the creation of real-time, fuel-efficient driving strategies, ultimately contributing to more sustainable transportation solutions.

2 Literature review

Recent advances in artificial intelligence and machine learning have enabled researchers to develop accurate and reliable models for predicting fuel consumption for vehicles. For example, Topić et al. [14] used linear regression models and neural networks to predict fuel consumption based on vehicle velocity, acceleration, and road slope time series inputs. The results showed that the proposed neural network-based approach was accurate, fast, and suitable for a variety of applications, such as vehicle routing optimization, driving cycle validation, and transport planning. Also, Li et al. [10] predicted the fuel consumption of a 13,000 TEU class container ship using in-service data collected based on two methods, an artificial neural network (ANN) and multiple linear regression (MLR), to select the input variables and prevent overfitting and multicollinearity. The models based on ANN showed the best prediction accuracy with a goodness of fit between 0.9709 and 0.9936. Additionally, sensitivity analysis indicated an optimal draught of 14.79 m for the ship, providing optimal fuel consumption efficiency. The models could be of great use for ship operators to make decisions for efficient operation. Moreover, Hamed et al. [15] proposed a machine-learning model based on the support vector machine algorithm to predict vehicle fuel consumption. The proposed model was tested on an on-board diagnostics dataset using 18 features and achieved an R-squared metric value of 0.97, which is higher than other related work using the same algorithm. The model could help manufacturers find successful Fuel Consumption Prediction models, thus improving business economics and satisfying domain needs. Furthermore, Alamdari et al. [16] used machine learning techniques to model and predict the fuel consumption of haul trucks in an open pit mine. Multiple linear regression, random forest, artificial neural network, support vector machine, and kernel nearest neighbor were tested on an actual dataset from an Iron ore open pit mine in Iran. The artificial neural network showed the best performance, with a coefficient of determination of 0.903, mean square error of 489.173, and mean absolute error of 13.440. A sensitivity analysis was also conducted to evaluate the significance of the independent variables.

Additionally, Perrotta et al. [7] used three Machine Learning techniques (SVM, RF, and ANN) to model the fuel consumption of articulated trucks using telematic data from standard sensors, as well as road characteristics from the highways agency pavement management system. Results of the study show that all three methods can be used to develop models with good accuracy, with the random forest model having the best performance. The model could be used to help fleet managers review existing vehicle routing decisions and road managers to better understand the fuel consumption of vehicles and how it is affected by road geometry. Besides, Yao et al. [2] discovered the correlation between driving behavior and fuel consumption by matching data collected from mobile phones and onboard diagnostic systems (OBD) in taxis. Three prediction models (backpropagation (BP) neural network, support vector regression (SVR), and random forests) were used to predict fuel consumption based on mobile phone data. The results show that speed, acceleration, deceleration, and cruising time are important fuel consumption indicators, and all three models had a relative error of less than 10%. Recent studies, such as those by Suarez et al. [17], Zhao et al. [18] and Pereira et al. [19] have highlighted the effectiveness of machine learning approaches, such as SVM and Random Forest, in fuel consumption prediction. These models are highly accurate and efficient, making them relevant for real-time fuel efficiency monitoring in transportation systems.

Other studies focused on evaluating fuel efficiency. For example, Hao et al. [20] investigated the driving behavior evaluation model to evaluate fuel efficiency using data from five trucks and extracted four characteristic parameters. They used K-means clustering combined with a density-based clustering non-parametric algorithm (DBSCAN) to cluster the parameters into three types of driving behavior, which were labeled as low, medium, and high fuel consumption. The model was trained on data and was found to have a prediction accuracy rate of 77.13%. Moreover, Ping et al. [21] introduced two machine-learning methods for evaluating the fuel efficiency of driving behavior using naturalistic driving data. A model is developed through spectral clustering and deep learning-based object detection to predict the fuel consumption associated with different driving behaviors. This model can then be applied in advanced driving assistance systems to improve vehicle fuel economy, decrease the energy cost of transportation, and develop better behavior assessment technology. We also acknowledge the inherent non-linearity in the relationship between acceleration and speed, as highlighted in previous works [22, 23]. These studies demonstrate that non-linear models can capture more complex driver behaviors and powertrain dynamics.

3 Methods

3.1 Proposed framework

The diagram shown in Fig. 1 presents a comprehensive framework for analyzing the relationship between driver behavior and fuel consumption using a combination of data collection, modeling, and machine learning techniques. This framework is designed to incorporate driving behavior into fuel consumption models and extract meaningful insights that correlate driving patterns with fuel efficiency, ultimately aiming to inform better driving practices and enhance fuel economy.

The theoretical construct of the proposed method is centered around integrating driving behavior into fuel consumption modeling, which entails employing a multi-layered approach. This framework is based on the premise that driving patterns significantly impact vehicle fuel efficiency and that understanding these patterns at both macro and micro levels can yield actionable insights for improving fuel economy. To achieve this, the framework uses a combination of

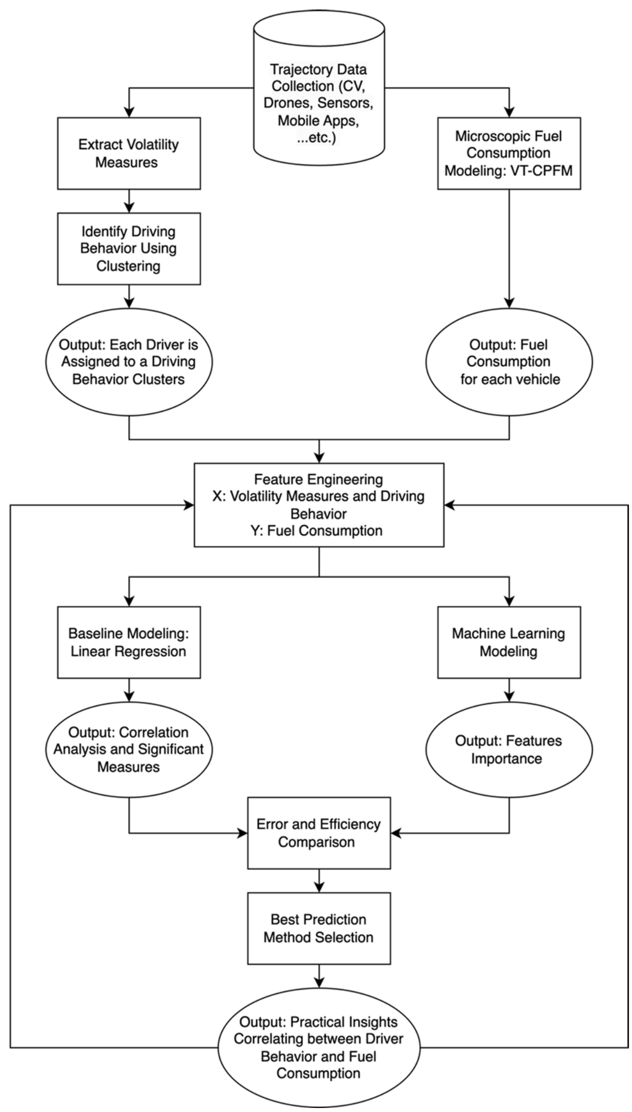

Fig. 1 A comprehensive framework for analyzing the relationship between driver behavior and fuel consumption

real-world data analysis, machine learning, and physics-informed models, which forms a comprehensive system that bridges the gap between driver behavior and fuel consumption.

At its core, the framework starts by defining different driving behaviors—such as conservative, normal, and aggressive—and linking these behaviors to corresponding fuel consumption rates. By analyzing extensive naturalistic driving data, such as the dataset of more than 13,500 road users in a mixed traffic environment, the method captures a wide variety of driving styles and traffic conditions. The first step involves establishing a baseline correlation between driving volatility measures (e.g., speed variations, accelerations, and decelerations) and fuel consumption through linear regression modeling. This statistical model serves as an initial indicator of how different driving behaviors affect fuel usage but is limited by its moderate accuracy.

This framework's core is the Trajectory Data Collection phase, which aggregates data from various sources such as connected vehicles (CV), drones, sensors, and mobile applications. These sources collectively provide rich, high-resolution data on vehicle trajectories, capturing complex and instantaneous details of driving behavior. The next phase involves the Microscopic Fuel Consumption Modeling, specifically utilizing the VT-CPFM (Virginia Tech Comprehensive Power-based Fuel Consumption Model). This model is employed to estimate fuel consumption at a granular level for each vehicle based on the collected trajectory data. The output of this phase is a detailed quantification of fuel consumption for each vehicle under study.

Following the data collection and initial fuel consumption modeling, the framework moves into the analysis phase, where Volatility Measures are extracted. These measures are crucial for characterizing the variability and aggressiveness in driving behavior, encompassing metrics such as acceleration, deceleration, and speed variations. With these measures in hand, clustering techniques are applied to identify three distinct driving behavior clusters, including normal, conservative, and aggressive. This clustering assigns each driver to specific behavior clusters, enabling a structured analysis of different driving styles.

The framework then proceeds to Feature Engineering, where the extracted volatility measures and identified driving behavior clusters (denoted as X) are paired with the corresponding fuel consumption data (denoted as Y). This step transforms the raw data into a structured format suitable for further analysis. Linear regression modeling is employed to establish baseline correlations. This baseline model helps in understanding the fundamental relationships between the features (volatility measures and driving behavior) and the target variable (fuel consumption). The output from this stage includes key correlation metrics and the identification of significant measures influencing fuel consumption. Parallel to the baseline modeling, advanced Machine Learning Modeling techniques are applied. These techniques are designed to capture more complex, non-linear relationships between the driving behavior features and fuel consumption. The output from this stage highlights the importance of various features, providing deeper insights into the factors that most significantly impact fuel efficiency. The framework concludes with an Error and Efficiency Comparison between the baseline linear regression model and the machine learning models. This comparison is crucial for determining the best predictive method, balancing accuracy and computational efficiency. Finally, the best prediction method is selected, and the framework generates practical insights. These insights correlate specific driving behaviors with fuel consumption patterns, offering actionable recommendations for drivers to optimize their driving habits and reduce fuel usage, which could be used as feedback for the feature engineering step.

Using driving behavior to predict fuel consumption is crucial because it directly influences the efficiency of vehicle operation. Factors such as acceleration, deceleration, and speed variability play significant roles in determining how much fuel a vehicle consumes. By understanding and modeling these behaviors, predictions can become more accurate and tailored to individual driving patterns. The advantages of this framework include its comprehensive data collection from diverse sources, detailed microscopic fuel consumption modeling, and the application of both traditional statistical methods and advanced machine learning techniques. This holistic approach ensures robust, precise predictions and offers actionable insights for drivers to improve their habits, leading to enhanced fuel efficiency, cost savings, and reduced environmental impact.

3.2 Dataset

The dataset includes about 5,408 drivers and was collected and reported in Germany. The dataset was collected in a mixed traffic environment and urban settings that involves cars, trucks, and bus users, as well as pedestrians and cyclists, at four intersections in 10 h of measurement and is called inD data (intersection Drone dataset) [24]. The data covers, for instance, the recording locations, times, and lengths along with the type of road user, the length of the track, or the average speed of each vehicle. The extracted trajectories in inD dataset include the x and y coordinations of the road user, width of the vehicle, length of the vehicle, the velocity in the x and y directions, the acceleration in x and y directions, the velocity of a road user as it moves in a forward direction (i.e., longitudinal), speed of a lateral movement of a vehicle as it transitions through a turn or travels along a curved path, the component of the vector acceleration of a point in the vehicle in the x direction, the acceleration coming when car is rotating around the corner center, and the distance between a vehicle and the vehicle in front of it (i.e., heading).

3.3 Driver behavior classification model

We adopted a framework to classify the driving behavior. This was previously developed and tested in several studies [25–28]. In this research, we use the same main components of the framework. The first step is to extract features from the trajectory data of each driver using volatility measures, which have been used in the literature [25, 29]. This is to identify the behavior of drivers as a significant safety parameter. Higher volatility measures were also observed to indicate a greater likelihood of the driver being unstable and risky, suggesting increased aggressiveness [29, 30]. We used ten different volatility measures (VM_i), as shown in Table 1. These measures provide insights into the typical driving behaviors and their variabilities, which are crucial for developing predictive models of fuel consumption. Through the analysis of these metrics, the framework can link particular driving behaviors with fuel consumption, enabling more precise predictions and practical strategies to enhance fuel efficiency. Second, we use these extracted features as inputs for an unsupervised machine-learning algorithm to cluster the behavior of each driver into normal, conservative, and aggressive. Conservative drivers are drivers who typically drive at a lower speed than the average speed of other vehicles. They exhibit a higher possibility of making slower and more cautious lane-changing maneuvers. Normal drivers are drivers who generally adhere to speed limits and follow typical driving patterns. When required to change lanes, they do so smoothly without abrupt acceleration or deceleration, ensuring the maneuver is completed safely without disrupting the flow of traffic. Aggressive drivers typically exceed average speed limits and engage in risky maneuvers like overtaking and abrupt lane

Table 1 Volatility measures (where V: speed, D_{long}: longitudinal deceleration, A_{long}: longitudinal acceleration, and AD_{long} : longitudinal deceleration or acceleration)

Volatility measure	Description	Equation
$\boldsymbol{VM_1}$	Standard deviation of speed	$\sqrt{\frac{\sum_{i=1}^{N}\left(V_i-\bar{V}\right)^2}{N}}$
$\boldsymbol{VM_2}$	Standard deviation of longitudinal deceleration or acceleration	$\sqrt{\frac{\sum_{i=1}^{N}\left(AD_{long_i}-\bar{AD}_{long}\right)^2}{N}}$
$\boldsymbol{VM_3}$	Coefficient of variation of speed	$100\times\frac{\sqrt{\frac{\sum_{i=1}^{N}(V_i-V)^2}{N}}}{V}$
$\boldsymbol{VM_4}$	Coefficient of variation of longitudinal acceleration	$100\times\frac{\sqrt{\frac{\sum_{i=1}^{N}\left(A_{long_i}-A_{long}\right)^2}{N}}}{\bar{A}_{long}}$
$\boldsymbol{VM_5}$	Coefficient of variation of longitudinal deceleration	$100\times\frac{\sqrt{\frac{\sum_{i=1}^{N}\left(D_{long_i}-D_{long}\right)^2}{N}}}{\bar{D}_{long}}$
$\boldsymbol{VM_6}$	Mean absolute deviation of speed	$\frac{\sum_{i=1}^{N}\left\|V_i-\bar{V}\right\|}{N}$
$\boldsymbol{VM_7}$	Mean absolute deviation of longitudinal acceleration	$\frac{\sum_{i=1}^{N}\left\|A_{long_i}-\bar{A}_{long}\right\|}{N}$
$\boldsymbol{VM_8}$	Quantile coefficient of variation of normalised speed	$100\times\frac{Q_{V3}-Q_{V1}}{Q_{V3}+Q_{V1}}$, where Q_1 and Q_3 are the sample 25^{th} and 75^{th} percentiles.
$\boldsymbol{VM_9}$	Quantile coefficient of variation of longitudinal acceleration	$100\times\frac{Q_{A\,long_3}-Q_{A\,long_1}}{Q_{A\,long_3}+Q_{A\,long_1}}$
$\boldsymbol{VM_{10}}$	Quantile coefficient of variation of longitudinal deceleration	$100\times\frac{Q_{D\,long_3}-Q_{D\,long_1}}{Q_{D\,long_3}+Q_{D\,long_1}}$

changes. They often choose to accelerate rather than slow down or stop, favoring speed over safety. In this study, we will use K-means algorithm, which was used successfully in previous studies [25–28].

3.4 Fuel consumption model

We used VT-CPFM as a microscopic fuel consumption model that depends on instantaneous vehicle power [31]. This model was designed to eliminate the necessity of calibrating parameters through field tests or simulators and to prevent the occurrence of the bang-bang control effect. It can avoid the bang-bang control as it has a second-degree polynomial relationship with vehicle-specific power (VSP). Furthermore, the model can be applied to publicly available data, which makes it feasible to be used in any geographical location. This acknowledges the inherent non-linearity in the relationship between acceleration and speed, as highlighted in previous works [22, 23]. The model was also tested for different vehicle types, including light- and heavy-duty vehicles and buses [32, 33]. More details on the model and its implementation can be found in Rakha et al. [31].

First, power can be calculated using the following Eq. (1):

$$P(t_i) = \left(R(t_i) + \frac{1.04ma(t_i)}{3{,}600\eta_d}\right) * v(t_i) \tag{1}$$

where $P(t_i)$ = power at time step t_i (kW), m = vehicle mass (kg), $a(t_i)$ = vehicle acceleration at time step t_i (m/s^2), $v(t_i)$ = vehicle speed at time step t_i (km/h), η_d = parameter of driveline efficiency, and $R(t_i)$ = force of resistance at time step t_i (N).

$R(t_i)$ can be computed using the following Eq. (2):

$$R(t_i) = \frac{\rho}{25.92} C_D C_h A_f v(t_i)^2 + 9.8066m\frac{C_r}{1{,}000}\left(c_1 v(t_i) + c_2\right) + 9.8066mG(t_i) \tag{2}$$

where ρ = air density (1.2256 kg/m^3 at sea level and 15 °C), C_D= coefficient of vehicle drag (unitless), C_h = factor to correct for elevation, A_f = area of vehicle frontal (m^2), $G(t_i)$ = grade of roadway at time step t_i, and c_r, c_1, and c_2 = parameters for rolling resistance (unitless). The drag coefficient values used for the energy consumption estimation vary depending on vehicle type and aerodynamic properties. For this study, the drag coefficient values were calibrated based on literature and specific data from the dataset. For vehicles in the dataset, the drag coefficient ranged from 0.30 to 0.35 for light-duty vehicles and from 0.58 to 0.78 for heavy-duty, which is typical in urban driving conditions [31, 32, 34, 35].

Fuel consumption (*FC*) is then computed applying Eq. (3). More details on the needed parameters and their values can be found in [31].

$$FC(t_i) = \begin{cases} \alpha_0 + \alpha_1 P(t_i) + \alpha_2 P(t_i)^2 & \forall P(t_i) \geq 0 \\ \alpha_0 & \forall P(t_i) < 0 \end{cases} \tag{3}$$

where α_0, α_1, and α_2 are constants to account for vehicle-specific calibration in the model.

3.5 Baseline prediction model

We employ Multivariate Linear Regression modeling to establish baseline correlations between driving behavior and fuel consumption. This statistical technique is an extension of simple linear regression and is used to model the relationship between two or more explanatory (independent; X) variables and a single response (dependent, Y) variable [36]. In this study, the independent variables include various volatility measures and driving behavior metrics, while the dependent variable is fuel consumption. Multivariate Linear Regression helps in understanding the fundamental relationships between multiple features and their combined impact on fuel consumption. By fitting a linear equation to observed data, this model quantifies how changes in each of the independent variables influence the dependent variable, providing a comprehensive analysis of driving behavior's effect on fuel consumption. The general form of the multivariate linear regression is shown in Eq. (4) [36]:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + cdots + \beta_n X_n + \epsilon \tag{4}$$

where Y is the dependent variable (fuel consumption), β_0 is the intercept term, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients corresponding to each independent variable, $X_1, X_2, \ldots, X_n$ are the independent variables (volatility measures and driving behavior metrics), and ϵ is the error term. By applying this model, we can derive key correlation metrics, such as the coefficients β, which indicate the strength and direction of the relationship between each independent and dependent variable. Additionally, the model helps identify significant measures influencing fuel consumption, enabling us to pinpoint which aspects of driving behavior are most critical in determining fuel efficiency. The output from the multivariate linear regression model includes the coefficients, which indicate the impact of each independent variable on fuel consumption; p-values, which assess the statistical significance of each coefficient; R-squared, which is a measure of how well the model explains the variability in the dependent variable, and Mean Squared Error (MSE), which is a measure of the average squared difference between the actual values and the values predicted by the model. It quantifies the model's overall fit to the data; a lower MSE indicates a better fit [36].

This foundational analysis using multivariate linear regression provides a critical understanding of the interactions between driving behavior and fuel consumption, forming the basis for more advanced predictive modeling and insights into fuel-efficient driving strategies. While linear regression was used as a baseline model for its simplicity and interpretability, we acknowledge that more advanced models, such as Machine Learning algorithms, better capture the dynamic and non-linear nature of driving behavior. We also used linear regression because we wanted to build on its results an informed machine learning algorithm.

3.6 Machine learning model

Parallel to the baseline modeling, we applied Random Forest to capture more complex, non-linear relationships between the driving behavior features and fuel consumption. Random Forest is a powerful machine learning algorithm commonly used for both regression and classification tasks [36]. It belongs to the ensemble learning methods, which combine multiple individual models to improve predictive performance. Random Forest consists of a collection of decision trees, where each tree is built using a subset of the data and a subset of the features. At each split in the decision tree, Random Forest randomly selects a subset of features to consider, which helps in reducing correlation between trees and preventing overfitting. For regression tasks, the predictions of all the trees are averaged to obtain the final prediction. For classification tasks, a voting mechanism is used to determine the class label [36].

Random Forest is known for its robustness, scalability, and ability to handle high-dimensional datasets with many features [36–39]. It is also effective in capturing complex relationships between features and target variables, making it suitable for various predictive modeling tasks, including the analysis of fuel consumption patterns and the identification of important driving behavior features. Random Forest has been applied in many civil engineering applications and has proved to be promising as predictive model [40–45]. The output from this stage is a real-time predictive model and highlights the importance of various features, providing deeper insights into the factors that most significantly impact fuel efficiency. It is evaluated using R-squared and MSE.

4 Analysis and results

4.1 Driver behavior results

Using the driver behavior classification model, drivers were categorized into Normal, Aggressive, and Conservative clusters based on speed and acceleration variability and other volatility measures, Table 2 shows the number of drivers in each cluster.

Table 2 Results of driver behavior classification model

Cluster	Number of drivers
Normal	2555
Aggressive	1133
Conservative	1720

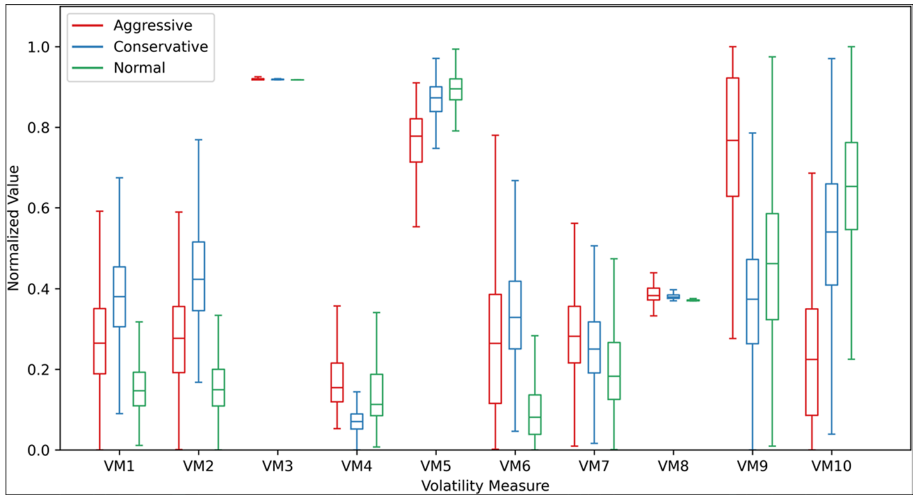

Fig. 2 Volatility measure for the different drivers clustering

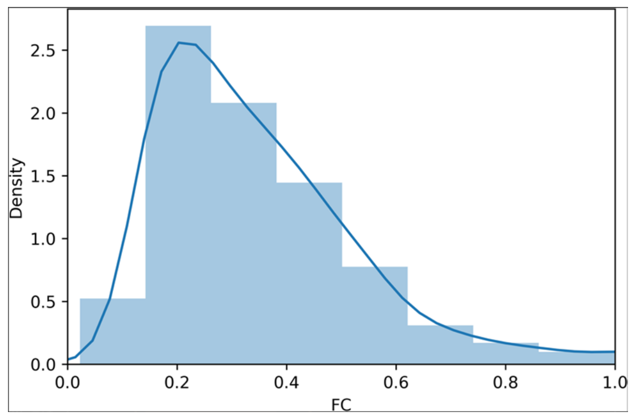

Fig. 3 Fuel consumption results

The Normal Cluster includes 2555 drivers, representing the most common driving style. These drivers maintain moderate speed and acceleration volatility values, resulting in stable, predictable driving behavior. They have low variability in speed and acceleration, leading to consistent driving patterns that avoid abrupt changes. The Aggressive Cluster consists of 1133 drivers characterized by high volatility in speed and acceleration. Their driving style features frequent and substantial changes in speed, marked by rapid acceleration and braking. The Conservative Cluster includes 1720 drivers who exhibit cautious driving behaviors. They have higher variability in speed and acceleration and prefer slower, more deliberate driving with smooth, gradual changes. These drivers tend to avoid high speeds and rapid maneuvers, optimizing fuel efficiency and enhancing safety by minimizing abrupt actions. The box plot in Fig. 2 illustrates the driving behaviors of the three driver clusters across a series of volatility measures (VM1 through VM10). Our analysis indicates that VM6 (Mean Absolute Deviation of Speed) and VM3 (Coefficient of Variation of Speed) had the most significant impact on fuel consumption. These measures capture variability in driving patterns, with higher deviations correlating to increased fuel usage, particularly in urban environments. These measures capture different aspects of driving volatility, including speed and acceleration variability.

4.2 Fuel consumption results

Using the VT-CPFM, we found the fuel consumption of each driver. The histogram and density plot in Fig. 3 offers valuable insights into the fuel consumption (FC) rate distribution among drivers, shedding light on how different driving behaviors influence fuel efficiency. Since the pattern is right-skewed, most drivers exhibit relatively low fuel consumption rates. The peak of the density plot is around an FC rate of 0.2, where the majority of drivers achieved moderate fuel efficiency.

The distribution highlights key driving behaviors: low FC rates (0.0 to 0.3) indicate fuel-efficient practices with steady speeds and minimal accelerations/decelerations. Moderate FC rates (0.3 to 0.6) suggest a mix of driving styles, occasionally less fuel-efficient. High FC rates (0.6 to 1.0) indicate aggressive driving with frequent accelerations and braking, prioritizing speed over fuel economy. We compared the resulted FC between different driving behavior. The box plot in Fig. 4 illustrates the FC across the three categories of driver behavior. Normal drivers exhibit the lowest and most consistent FC rates, with a median around 0.2 and an interquartile range (IQR) between 0.1 and 0.6. Conservative and aggressive drivers show a higher median FC rate of approximately 0.4 and a broader IQR from 0.1 to 0.9, reflecting a mix of driving behaviors. Results reveal that both conservative and aggressive driving behaviors lead to higher and more variable fuel consumption compared to normal driving. Normal drivers consistently maintain lower FC rates, indicating that smoother and more controlled driving behaviors are crucial for fuel efficiency. In contrast, the broader range and higher median FC rates for conservative and aggressive drivers highlight the inefficiencies and greater fuel demands associated with these driving styles. The higher mean fuel consumption for conservative drivers may be attributed to prolonged driving times at lower speeds, resulting in inefficiencies in fuel usage compared to more dynamic driving behaviors. Thus, promoting normal driving behaviors can significantly improve overall fuel economy and reduce variability in fuel consumption. The consistency in normal driving, characterized by adherence to speed limits and smooth lane changes, is the most fuel-efficient driving behavior. Despite their cautious driving, which includes slower speeds and more deliberate lane changes, the variability and occasional inefficiencies contribute to a higher and more variable FC rate in conservative driving. However, the aggressive driving style, involving high speeds and frequent aggressive maneuvers, leads to significantly higher fuel consumption and greater variability in FC rates.

4.3 Baseline prediction model

The baseline FC prediction model using Linear regression showed a reasonable fit, with an R-squared value of 0.511. This indicates that the model can explain approximately 51.1% of the variance in fuel consumption. The results are presented in Table 3. The mean squared error (MSE) of 0.031 reflects good prediction accuracy for the dataset. The variables of VM1 (Standard deviation of speed), VM2 (Standard deviation of longitudinal deceleration or acceleration), VM5 (Coefficient of variation of longitudinal deceleration), and VM9 (Quantile coefficient of variation of longitudinal acceleration) have negative coefficients of − 0.136, − 0.355, − 1.242, and − 0.423 respectively. This implies that higher values of these variables are associated with reduced fuel consumption. On the other hand, VM3, VM4, VM6, VM7, VM8, and VM10 positively impact fuel consumption with various coefficients indicating a substantial increase in FC for higher values of these variables. The coefficient of the normal cluster is − 0.001, with a p-value of 0.904, indicating an insignificant impact on fuel consumption. This suggests that normal driving behavior does not notably alter the FC from the baseline. However, the conservative cluster exhibits a small but significant negative coefficient of 0.021, indicating an increase in fuel consumption for conservative driving behaviors.

4.4 Machine learning prediction model

The machine learning prediction using Radom Forest model for FC has achieved outstanding performance, as indicated by an R-squared value of 0.956 and a mean squared error (MSE) of 0.003. These metrics demonstrate the model's strong predictive capabilities, with the R-squared value showing that the model accounts for 95.6% of the variance in fuel consumption. This high level of explained variance suggests that the model is highly accurate in predicting fuel

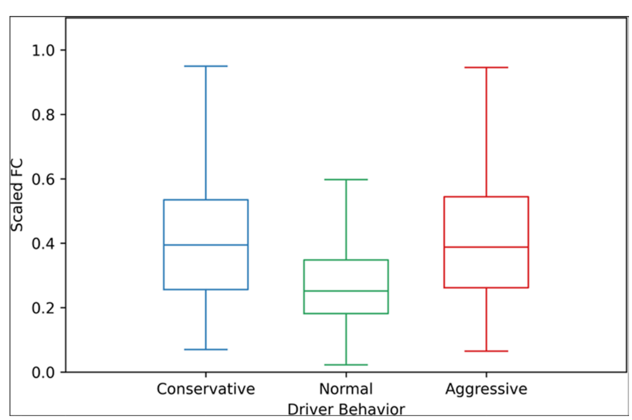

Fig. 4 Scaled fuel consumption rates

Table 3 Prediction model coefficients

Variable	Coefficient	*P*-value
Constant	0.218	< 0.05
VM1	− 0.136	< 0.05
VM2	− 0.355	< 0.05
VM3	2.034	< 0.05
VM4	0.763	< 0.05
VM5	− 1.242	< 0.05
VM6	1.101	< 0.05
VM7	0.542	< 0.05
VM8	2.284	< 0.05
VM9	− 0.423	< 0.05
VM10	0.272	< 0.05
Normal Cluster	− 0.001	0.904
Conservative Cluster	0.021	< 0.05

consumption, with minimal errors, as reflected in the low MSE. The feature importance analysis, depicted in the bar chart in Fig. 5, highlights the significant contributions of various numerical features to the model's accuracy. VM6 emerges as the most influential feature, causing the greatest decrease in model accuracy when omitted, suggesting its critical role in predicting fuel consumption. Other features, such as VM3, VM7, and VM8, also show considerable impacts, though to a lesser extent than VM6. These features collectively drive the model's predictive strength, while VM1, VM2, VM4, VM5, VM9, and VM10 have relatively lower influence on model performance. When we compare linear regression to the random forest technique, we observe some similarities, such as the high effect of the VM3, VM4, VM6, VM7, and VM8 variables.

4.5 Evaluation and comparison

Predictive modeling, like linear regression and Random Forest, provides insights into performance. Linear regression yields an R-squared of 0.511 and an MSE of 0.031, indicating moderate predictive accuracy. In contrast, Random Forest, an ensemble technique, achieves a higher R-squared of 0.956 and a lower MSE of 0.003, demonstrating superior predictive precision by capturing 95.6% of the variance in the dependent variable. Figure 6 shows that Random Forest outperforms Linear Regression in R-squared and MSE metrics. Random Forest predictions align closer to the ideal line = compared to Linear Regression, indicating superior accuracy in capturing dataset patterns. The skew in Fig. 6 is due to the driving behavior distribution in the dataset. A significant proportion of drivers exhibited aggressive and conservative behavior, which led to higher fuel consumption predictions. While both models generally fit the data well, Random Forest excels in predictive precision. Choosing between linear regression and Random Forest depends on project requirements. Random Forest's strong predictive capabilities are ideal for tasks demanding high accuracy. Conversely, linear regression remains suitable for scenarios valuing interpretability and simplicity over absolute predictive performance.

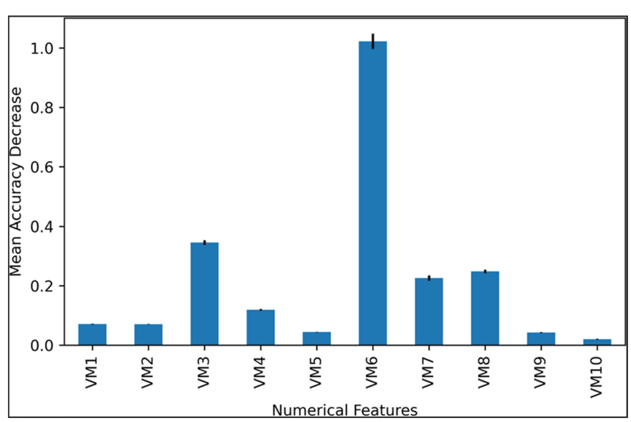

Fig. 5 Random forest model results

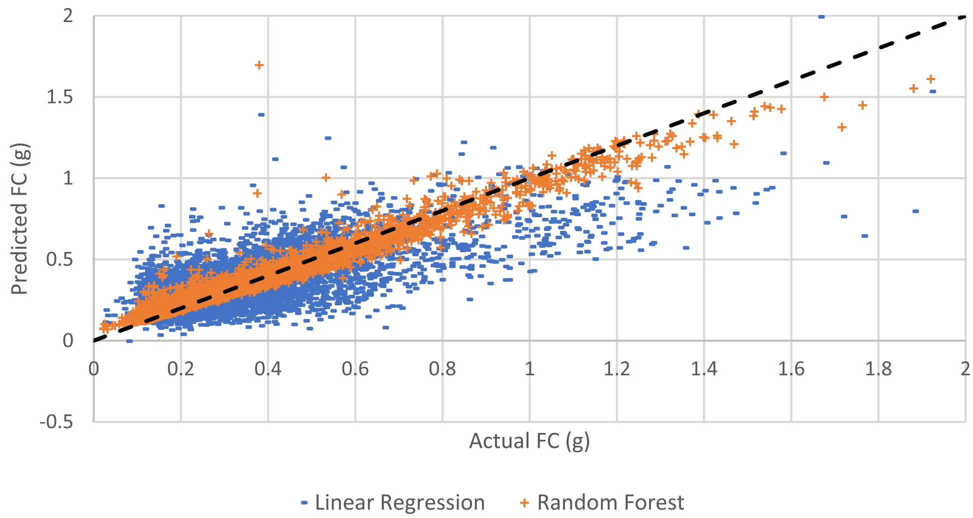

Fig. 6 Comparison between the linear regression and random forest performance

The choice between linear regression and Random Forest depends largely on the specific requirements of the project at hand and policymakers. Random Forest's superior predictive capabilities make it ideal for tasks that demand high accuracy and precision. This model is particularly well-suited for complex datasets where capturing non-linear relationships and interactions between variables is crucial. However, linear regression remains valuable in scenarios where interpretability and simplicity are prioritized over absolute predictive performance. Linear regression models are easier to understand and explain, making them useful for applications where transparency and straightforward analysis of the relationships between variables are important. For example, for Advanced Driver Assistance Systems (ADAS) and real-time fuel efficiency optimization, Random Forest, due to its ability to provide highly accurate predictions, crucial for real-time decision-making and precise fuel consumption monitoring. On the other hand, for preliminary analysis, network evaluation, or situations where stakeholders require clear and interpretable results, Linear regression can be chosen for its straightforward nature and ease of communication.

Recent studies have explored various eco-driving strategies [46], providing a strong foundation for the integration of real-time fuel consumption prediction models, such as ours, into ADAS. For instance, eco-driving for Fuel Cell Hybrid Electric Vehicles (FCHEVs) through signalized intersections has been approached as a coupled problem of speed planning and powertrain control under complex environmental constraints [47]. The insights gained from our model can be directly applied in ADAS to promote energy-efficient driving. ADAS systems rely on real-time monitoring and prediction of driver behavior, providing instant feedback to guide more efficient driving styles. By incorporating predictions from our Random Forest model, which accurately captures driving behavior's impact on fuel consumption, ADAS can help drivers adopt ecological driving practices. These include smoother acceleration, maintaining optimal speeds, and reducing unnecessary idling, all of which contribute to lower fuel consumption. Several recent studies, including the works of Chada et al. [48], Shi et al. [49], and Bakiballah et al. [50], have demonstrated that driver-assistance technologies can significantly reduce fuel consumption by promoting energy-efficient driving behaviors in real-time.

5 Conclusion

Driving behavior plays a crucial role in determining vehicle fuel consumption. Investigating the link between driving behavior and fuel usage can help lower transportation energy costs and advance the technology for behavior assessment in Advanced Driver Assistance Systems (ADAS). Understanding this relationship is essential for creating more environmentally friendly driving assistance systems and enhancing vehicle fuel efficiency. However, modeling driving behavior under dynamic conditions is complex, especially in mixed traffic environments and urban settings, where varying driver behaviors interact with each other and with vulnerable road users. This complexity makes quantitatively analyzing the connection between driving behavior and fuel consumption challenging.

The framework we propose serves both as a theoretical model and a practical tool, providing a comprehensive, multi-step process to correlate driving behavior with fuel consumption. This approach integrates traditional statistical methods with advanced machine learning techniques to extract actionable insights. For validation, we utilized naturalistic data comprising approximately 5,408 distinct road users navigating mixed traffic environments and urban settings in Germany. We employed a microscopic fuel consumption model for framework calibration and an unsupervised clustering algorithm to classify driver behavior. The framework begins with the development of a baseline linear regression model, which produces an R-squared value of 0.511 and an MSE of 0.031, indicating moderate predictive accuracy. We transitioned to using Random Forest, a more sophisticated model that achieved a higher R-squared value of 0.956 and a lower MSE of 0.003. Our analysis also reveals that both conservative and aggressive driving behaviors result in significantly higher and more variable fuel consumption compared to normal driving. Normal drivers consistently maintain lower FC rates, indicating that smoother and more controlled driving behaviors are crucial for fuel efficiency. In contrast, the broader range and higher median FC rates for conservative and aggressive drivers highlight the inefficiencies and greater fuel demands associated with these driving styles. Thus, promoting normal driving behaviors can significantly improve overall fuel economy and reduce variability in fuel consumption.

While both linear regression and Random Forest models fit the data well, Random Forest excels in predictive precision and is better suited for applications requiring high accuracy. Linear regression, on the other hand, is more appropriate for scenarios where simplicity and interpretability are more important than the absolute level of predictive performance. These insights help inform the choice of modeling approach based on the specific needs and goals of the project. Our findings indicate that the proposed method effectively uncovers the relationship between driving behavior and fuel consumption at both macro and micro levels. This approach enables comprehensive fuel consumption feature prediction, which can be applied in advanced driving assistance systems. Integrating physics-informed models, such as those for the vehicle powertrain and driver behavior, can enhance the generalization of machine learning approaches by embedding domain knowledge that cannot be easily captured from data alone. This hybrid modeling can offer better predictive accuracy, particularly when applied to diverse driving conditions and vehicle types. While our methodology reduces the need for storing detailed instantaneous data, the calculation of volatility measures still requires memory for storing time-series information, albeit in a compressed form. This method deepens the understanding of driver behavior and supports the development of real-time, fuel-efficient driving strategies, contributing to more sustainable transportation solutions.

Future research can include incorporating other fuel consumption models including incorporating for example a model for the vehicle powertrain and the driver, which could help towards the generalization of the machine learning approach. We will also test the use of cluster-agnostic algorithms (e.g. DBSCAN) for driving behavior modeling, which could increase the transferability of the model as well.

Author contributions Conceptualization, H.A., M.O., A.J., N.K., R.A., and M.E.; Methodology, H.A., M.O., N.K., and R.A.; Software, H.A.; Validation, H.A., M.O., N.K., and R.A.; Formal analysis, H.A., M.O., N.K., R.A., and M.E.; Investigation, H.A., M.O., A.J., N.K., R.A., and M.E.; Writing—original draft preparation, H.A., M.O., A.J., N.K., R.A., and M.E.; Writing—review and editing, H.A., M.O., A.J., N.K., R.A., and M.E.; Visualization, H.A., and A.J.; Supervision, H.A., and M.E.

Data availability The raw data supporting the conclusions of this article will be made available by the authors upon request.

Declarations

Competing interests The authors declare no competing interests.

References

1. Vyas J, Das D, Chaudhury S. DriveBFR: driver behavior and fuel-efficiency-based recommendation system. IEEE Trans Comput Soc Syst. 2021;9(5):1446–55.
2. Yao Y, et al. Vehicle fuel consumption prediction method based on driving behavior data collected from smartphones. J Adv Transp. 2020;2020:1–11.

3. Kedia RK, Naick BK. Review of vehicle route optimisation. In: 2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE), IEEE, 2017, pp. 57–61.
4. Hu L, et al. Optimal route algorithm considering traffic light and energy consumption. IEEE Access. 2018;6:59695–704.
5. Issa A, Zawawi Z, Ashqar HI. Impact of road grade on fuel consumption: potential savings in Nablus, Palestine. Case Stud Transp Policy. 2023;12:100973. https://doi.org/10.1016/j.cstp.2023.100973.
6. Ben-Chaim M, Shmerling E, Kuperman A. Analytic modeling of vehicle fuel consumption. Energies (Basel). 2013;6(1):117–27.
7. Perrotta F, Parry T, Neves LC. Application of machine learning for fuel consumption modelling of trucks, in 2017 IEEE International Conference on Big Data (Big Data), IEEE, 2017, pp. 3810–3815.
8. Suh B, Chang YH, Han SB, Chung YJ. Simulation of a powertrain system for the diesel hybrid electric bus. Int J Autom Technol. 2012;13:701–11.
9. Wang J, Rakha HA. Fuel consumption model for heavy duty diesel trucks: model development and testing. Transp Res D Transp Environ. 2017;55:127–41.
10. Kim Y-R, Jung M, Park J-B. Development of a fuel consumption prediction model based on machine learning using ship in-service data. J Mar Sci Eng. 2021;9(2):137.
11. Li Y, Tang G, Du J, Zhou N, Zhao Y, Wu T. Multilayer perceptron method to estimate real-world fuel consumption rate of light duty vehicles. IEEE Access. 2019;7:63395–402.
12. Lopez-Martin M, Sanchez-Esguevillas A, Hernandez-Callejo L, Arribas JI, Carro B. Novel data-driven models applied to short-term electric load forecasting. Appl Sci. 2021;11(12):5708.
13. Lopez-Martin M, Sanchez-Esguevillas A, Hernandez-Callejo L, Arribas JI, Carro B. Additive ensemble neural network with constrained weighted quantile loss for probabilistic electric-load forecasting. Sensors. 2021;21(9):2979.
14. Topić J, Škugor B, Deur J. Neural network-based prediction of vehicle fuel consumption based on driving cycle data. Sustainability. 2022;14(2):744.
15. Hamed MA, Khafagy MH, Badry RM. Fuel consumption prediction model using machine learning. Int J Adv Comput Sci Appl. 2021. https://doi.org/10.14569/IJACSA.2021.0121146.
16. Alamdari S, Basiri MH, Mousavi A, Soofastaei A. Application of machine learning techniques to predict haul truck fuel consumption in open-pit mines. J Min Environ. 2022;13(1):69–85.
17. Suarez J, Makridis M, Anesiadou A, Komnos D, Ciuffo B, Fontaras G. Benchmarking the driver acceleration impact on vehicle energy consumption and CO2 emissions. Transp Res D Transp Environ. 2022;107:103282. https://doi.org/10.1016/j.trd.2022.103282.
18. Zhao D, et al. A review of the data-driven prediction method of vehicle fuel consumption. Energies. 2023;16:5258. https://doi.org/10.3390/en16145258.
19. Pereira G, Parente M, Moutinho J, Sampaio M. Fuel consumption prediction for construction trucks: a noninvasive approach using dedicated sensors and machine learning. Infrastruct. 2021;6(11):157. https://doi.org/10.3390/infrastructures6110157.
20. Hao R, Yang H, Zhou Z. Driving behavior evaluation model base on big data from internet of vehicles. Int J Ambient Comput Intell (IJACI). 2019;10(4):78–95.
21. Ping P, Qin W, Xu Y, Miyajima C, Takeda K. Impact of driver behavior on fuel consumption: classification, evaluation and prediction using machine learning. IEEE Access. 2019;7:78515–32.
22. Makridis M, Fontaras G, Ciuffo B, Mattas K. MFC free-flow model: introducing vehicle dynamics in microsimulation. Transp Res Rec J Transp Res Board. 2019;2673(4):762–77. https://doi.org/10.1177/0361198119838515.
23. Makridis MA, Kouvelas A. Adaptive physics-informed trajectory reconstruction exploiting driver behavior and car dynamics. Sci Rep. 2019;13(1):1121. https://doi.org/10.1038/s41598-023-28202-1.
24. Bock J, Krajewski R, Moers T, Runde S, Vater L, Eckstein L. The ind dataset: a drone dataset of naturalistic road user trajectories at German intersections. In: 2020 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2020, 1929–34.
25. Khanfar NO, Elhenawy M, Ashqar HI, Hussain Q, Alhajyaseen WKM. Driving behavior classification at signalized intersections using vehicle kinematics: application of unsupervised machine learning. Int J Inj Contr Saf Promot. 2022;30:1–11.
26. Dwekat TM, Almsre AA, Ashqar HI. Effect of roundabout design on the behavior of road users: a case study of roundabouts with application of Unsupervised Machine Learning, arXiv preprint arXiv:2309.14540, 2023.
27. Khanfar NO, Ashqar HI, Elhenawy M, Hussain Q, Hasasneh A, Alhajyaseen WKM. Application of unsupervised machine learning classification for the analysis of driver behavior in work zones in the state of Qatar. Sustainability. 2022;14(22):15184.
28. Hamad FA, Hasiba R, Shahwan D, Ashqar HI. How Do drivers behave at roundabouts in a mixed traffic? A case study using machine learning. arXiv preprint arXiv:2309.13442. 2023.
29. Arvin R, Kamrani M, Khattak AJ. How instantaneous driving behavior contributes to crashes at intersections: extracting useful information from connected vehicle message data. Accid Anal Prev. 2019;127:118–33.
30. Mohammadnazar A, Arvin R, Khattak AJ. Classifying travelers' driving style using basic safety messages generated by connected vehicles: application of unsupervised machine learning. Transp Res Part C Emerg Technol. 2021;122:102917.
31. Rakha HA, Ahn K, Moran K, Saerens B, Van den Bulck E. Virginia tech comprehensive power-based fuel consumption model: model development and testing. Transp Res D Transp Environ. 2011;16(7):492–503.
32. Edwardes W, Rakha H. Virginia tech comprehensive power-based fuel consumption model: modeling diesel and hybrid buses. Transp Res Rec. 2014;2428(1):1–9.
33. Albool I, et al. Fuel consumption at signalized intersections: investigating the impact of different signal indication settings. Case Stud Transp Policy. 2023;13:101022.
34. Bansal R, Sharma RB. Drag reduction of passenger car using add-on devices. J Aerodyn. 2014;2014:1–13. https://doi.org/10.1155/2014/678518.
35. Paul AR, Jain A, Alam F. Drag reduction of a passenger car using flow control techniques. Int J Autom Technol. 2019;20(2):397–410. https://doi.org/10.1007/s12239-019-0039-2.
36. James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning, vol. 112. Berlin: Springer; 2013.
37. Ali J, Khan R, Ahmad N, Maqsood I. Random forests and decision trees. Int J Comput Sci Issues (IJCSI). 2012;9(5):272.

38. Jaber A, Csonka B. Temporal travel demand analysis of irregular bike-sharing users. In: International Conference on Human-Computer Interaction. Springer; 2022, pp. 517–525.
39. Jaber A, Al-Sahili K. Severity of pedestrian crashes in developing countries: analysis and comparisons using decision tree techniques. SAE Int J Transp Saf. 2022;11:307–20.
40. Ashqar HI, Shaheen QHQ, Ashur SA, Rakha HA. Impact of risk factors on work zone crashes using logistic models and Random Forest. In: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC), IEEE, 2021, pp. 1815–1820.
41. Ashqar HI, Elhenawy M, Almannaa MH, Ghanem A, Rakha HA, House L. Modeling bike availability in a bike-sharing system using machine learning. In: 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), IEEE, 2017, pp. 374–378.
42. Yuan H, Li G. A survey of traffic prediction: from spatio-temporal data to intelligent transportation. Data Sci Eng. 2021;6:63–85.
43. Ashqar HI, Elhenawy M, Rakha HA, Almannaa M, House L. Network and station-level bike-sharing system prediction: a San Francisco bay area case study. J Intell Transp Syst. 2021;26:1–11.
44. Iranitalab A, Khattak A. Comparison of four statistical and machine learning methods for crash severity prediction. Accid Anal Prev. 2017;108:27–36.
45. Ashqar HI, Elhenawy M, Rakha HA. Modeling bike counts in a bike-sharing system considering the effect of weather conditions. Case Stud Transp Policy. 2019;7(2):261–8.
46. Jayson T, Bakibillah ASM, Tan CP, Kamal MAS, Monn V, Imura J. Electric vehicle eco-driving strategy at signalized intersections based on optimal energy consumption. J Environ Manage. 2024;368:122245. https://doi.org/10.1016/j.jenvman.2024.122245
47. Liu B, Lu B, Sun C, Wang B, Jia B, Sun F. Convex optimization for long-term eco-driving of fuel cell hybrid electric vehicles on signalized corridors. IEEE Trans Veh Technol. 2024. https://doi.org/10.1109/TVT.2024.3443106.
48. Chada SK, Görges D, Ebert A, Teutsch R, Subramanya SP. Evaluation of the driving performance and user acceptance of a predictive eco-driving assistance system for electric vehicles. Transp Res Part C Emerg Technol. 2023;153:104193. https://doi.org/10.1016/j.trc.2023.104193
49. Shi X, Zhang J, Jiang X, Chen J, Hao W, Wang B. Learning eco-driving strategies from human driving trajectories. Physica A. 2024;633:129353. https://doi.org/10.1016/j.physa.2023.129353.
50. Bakibillah ASM, Kamal MAS, Tan CP, Hayakawa T, Imura J. Optimal eco-driving scheme for reducing energy consumption and carbon emissions on curved roads. Heliyon. 2024;10(1):e23586. https://doi.org/10.1016/j.heliyon.2023.e23586.