

Article

On the Use of Biofuels for Cleaner Cities: Assessing Vehicular Pollution through Digital Twins and Machine Learning Algorithms

Matheus Andrade ¹, Morsinaldo Medeiros ¹, Thaís Medeiros ¹, Mariana Azevedo ¹, Marianne Silva ¹, Daniel G. Costa ² and Ivanovitch Silva ^{1,*}

¹ UFRN-PPgEEC, Postgraduate Program in Electrical and Computer Engineering, Federal University of Rio Grande do Norte, Natal 59078-970, Brazil; matheus.diniz.122@ufrn.edu.br (M.A.); morsinaldo.medeiros.075@ufrn.edu.br (M.M.); thais.araujo.707@ufrn.edu.br (T.M.); mariana.brito.110@ufrn.edu.br (M.A.); marianne.silva@ufrn.br (M.S.)
² SYSTEC-ARISE, Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal; danielgcosta@fe.up.pt
* Correspondence: ivanovitch.silva@ufrn.br

Abstract: The air pollution caused by greenhouse gas emissions, particularly carbon dioxide (CO₂), is a significant environmental concern that impacts air quality and contributes to global warming. The transportation sector plays a pivotal role in this issue, being a major contributor to CO₂ emissions. In light of this situation, this article proposes a methodology that utilizes a supervised learning algorithm to estimate CO₂ emissions and compare vehicles fueled with ethanol and gasoline. Additionally, the solution adopts an online, unsupervised machine learning algorithm to identify data outliers and improve the confidence in the results. Furthermore, this work incorporates the concept of digital twins, using virtual models of vehicles to carry out more extensive pollution simulations and allowing the simulation of various types of vehicles and the modeling of realistic traffic scenarios. A supervised machine learning approach was adopted to infer emission data in the model, allowing more comprehensive and meaningful comparisons between real-world and simulated measurements. The performed analyses of pollution emissions for different speeds and sections of routes demonstrate that CO₂ emissions from ethanol were significantly lower than those from gasoline, favoring more sustainable fuels even in combustion engine vehicles. Adopting cleaner fuels is perceived as crucial to mitigate the negative effects of climate change, with plant-based fuels like ethanol being crucial during the transition from fossil fuels to a more sustainable vehicular landscape.

Keywords: machine learning; CO₂ emissions; vehicular pollution; digital twins; climate change mitigation; smart cities



Citation: Andrade, M.; Medeiros, M.; Medeiros, T.; Azevedo, M.; Silva, M.; Costa, D.G.; Silva, I. On the Use of Biofuels for Cleaner Cities: Assessing Vehicular Pollution Through Digital Twins and Machine Learning Algorithms. *Sustainability* **2024**, *16*, 708. <https://doi.org/10.3390/su16020708>

Academic Editors: Dan Liu, Zhongyi Zuo and Zhongzhen Yang

Received: 27 December 2023

Revised: 10 January 2024

Accepted: 11 January 2024

Published: 13 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Air pollution in urban areas has been intensifying in recent years, with direct implications for human health and the global ecosystem [1]. Carbon dioxide (CO₂) emissions from vehicles are one of the main contributors to this scenario, becoming a recurring topic in scientific, political, and social debates due to their impact on air quality [2]. For the expected transformations in the urban landscapes when dealing with the ongoing urbanization challenges and the urgent need for sustainable energies and resources, the pursuit of cleaner technologies and fuels will be one of the core concerns in this century [3–5].

In the context of greenhouse gases (GHGs), CO₂ is particularly significant, comprising 76% of total GHG emissions globally [6]. This underscores the critical role of the transportation sector, which is responsible for approximately 15% of global emissions [7]. Such statistics highlight the urgency of addressing CO₂ emissions within this sector, illustrating the potential impact of targeted mitigation strategies in reducing such numbers.

Given this challenging scenario, research on transportation carbon emissions has been widely considered in different regions and countries worldwide, receiving increasing attention [8]. Overall, it is essential to adopt technologies and policies that promote the reduction of CO₂ emissions and the use of renewable energies to combat climate change and improve air quality [2,9]. Hence, the global transition to clean-energy vehicles has gained prominence in reducing greenhouse gas emissions and decreasing the dependence on fossil fuels [10]. In this scenario, electric, hybrid, and biofuel-powered vehicles represent sustainable solutions to mitigate carbon emissions, standing out as crucial components in the energy matrix [11–14]. Additionally, achieving low-carbon mobility goals requires the implementation of strategies aimed at reducing vehicle emissions, including the promotion of renewable energy use, the improvement of vehicles' energy efficiency, and the installation of adequate infrastructure for electric and hybrid vehicles [9,15].

In attempts to deal with this stringent energy transformation challenge, the Internet of Things (IoT) has emerged as a potentially effective solution for monitoring and optimizing vehicle performance. By enabling the connection and intercommunication between smart objects, the IoT enables real-time data collection, providing valuable insights into vehicle operation [16,17]. In the automotive field, IoT solutions can be designed around On-Board Diagnostics (OBD-II), a tool that provides access to vehicle data, including information about CO₂ emissions. This resource facilitates continuous emission monitoring with the potential identification of areas for optimization and reduction, driving the transition towards sustainable mobility.

With the aim of fostering the development of improved policies for enhancing air quality and reducing the reliance on fossil fuels, this article introduces a methodology designed to generate valuable information. This information, in turn, supports the transition towards a more sustainable energy matrix within the transportation sector. The proposed methodology integrates the instrumentation between OBD-II and smartphones to capture real data from vehicle sensors. The retrieved data are used to indirectly compute CO₂ emissions through a developed estimation module, which also applies an unsupervised machine learning technique to remove data outliers that may be common in this type of monitoring [18,19]. Such an approach could even be adopted in the context of machine learning on low-power devices (TinyML), potentially enabling the application of machine learning models on resource-constrained devices, such as microcontrollers, for intelligent decision-making on the edge, which is expected to be one of the next revolutions in the automotive sector [20,21]. In this article, by adopting a smartphone-based approach with processing on the cloud, the developed solution may become more reproducible while also remaining highly adequate for embedding into vehicles, potentially contributing to the ongoing sustainable transformation process in this domain.

As an important step to stimulate even further the adoption of more sustainable fuels, potentially deepening our understanding and potential analyses of vehicle emissions, the concept of digital twins was also incorporated into our approach, providing virtual models that faithfully replicate real-world entities or processes [22,23]. In this article, the Simulation of Urban Mobility (SUMO) traffic simulator was exploited to create the intended digital twins and perform detailed pollution analyses [24]. These virtual models are accurate and reflect real vehicle behavior, offering an enhanced view of emissions and extending the achieved results for analysis.

By integrating SUMO, we expand our ability to assess environmental impacts by facilitating the comparisons among different types of fuels. However, a practical challenge emerges due to the type of data that is modeled by the tool, which is different from the data retrieved via the OBD-II interface. In this case, a supervised machine learning model that was trained with real data collected from vehicles was designed, allowing inferences about missing data and meaningful comparisons between both approaches. Thus, the integration of SUMO into our methodology allowed for a comprehensive understanding of emissions patterns and specific areas for targeted interventions.

Therefore, the contributions of this article are threefold:

- A practical approach to collecting data from vehicles through their OBD-II interfaces, which are retrieved through a smartphone and processed on the cloud via an unsupervised machine learning algorithm to remove outliers. The processed data are then used to indirectly estimate CO₂ emissions using a mathematical formulation;
- A digital twin approach based on the SUMO tool, allowing more extensive pollution assessment in a simulated environment. A supervised machine learning regression model was trained with previously collected data in order to allow the estimation of pollution emissions on a more realistic basis;
- Extensive comparisons of pollution emissions for vehicles fueled with gasoline and ethanol, for both real-world and simulation environments, enabling important discussions about the role of biofuels for sustainable transportation.

Since the energy transition is indispensable to achieving the Sustainable Development Goals established by the UN (United Nations), particularly Goal 13—Climate Action [25]—it is expected that the proposed approach can be a valuable contribution when reinforcing the need for a more urgent transition from fossil fuels to more sustainable alternatives [26].

The remainder of this paper is organized as follows. Section 2 presents related works that influenced our defined methodology and implementation. Section 3 provides details of the proposed method. Section 4 describes the conducted case study. Section 5 discusses the main obtained results, and finally, Section 6 presents conclusions and promising directions for future research.

2. Related Works

Several research works have investigated different approaches and methodologies to understand and quantify the environmental impact of the transportation sector. Some of them have also proposed effective strategies for emission reductions. These works have employed various techniques to collect data, influencing our research in multiple ways.

The work in [27] implemented an exhaust gas sensor positioned near a vehicle's exhaust system to enable the real-time monitoring and visualization of carbon monoxide (CO) and smoke emissions. Though promising, their approach had limitations, such as potential accuracy issues due to external factors and other gases present in the environment and the inability to differentiate emissions from different types of vehicles.

The authors of [28] proposed the use of OBD-II data transmitted to the cloud and the application of a long short-term memory (LSTM) model for efficient monitoring of CO₂ emissions. Such an approach, though practical in some contexts, required supervised training datasets, constraining its applicability.

The work in [29] utilized IoT dongles installed in vehicles for sensor readings, also applying an LSTM network to predict CO₂ emissions. Their system aimed to monitor vehicle emissions but faced the limitations of requiring a stable internet connection and limited data collection from only two vehicles in their experiments.

From a different perspective, ref. [30] used a TinyML model in an OBD-II automotive scanner to estimate CO₂ emissions. The proposed TinyML algorithm processed data using unsupervised learning, enabling the more accurate detection of noisy and outlier data. That approach enabled the low-cost monitoring of vehicle emissions through an embedded system approach, facilitating continuous monitoring, although only gasoline was considered as a fuel in that work.

Concerning simulations and virtual scenarios, several studies have strategically employed the SUMO tool, a versatile and widely adopted simulator renowned for its detailed and comprehensive analysis of urban traffic and mobility scenarios [24]. Leveraging SUMO's adaptability and robust simulation capabilities, many works have delved into intricate details, offering a nuanced understanding of the intersection between transportation, urban environments, and environmental sustainability [31]. This is due to the fact that this tool serves as a pivotal asset in meticulously exploring and dissecting the complexities associated with the challenging urban transportation scenario.

The authors of [29] intended to estimate air quality in diverse city areas, aiming to raise awareness and assist citizens in making informed decisions. Their proposal incorporated a traffic modeling approach that utilized historical traffic data, the SUMO traffic simulator, and a trajectory generation strategy to predict traffic volumes at different road segments and hours. Additionally, a pollution modeling approach employed the Vehicular Emissions INventories (VEIN) R package to estimate NO_x emissions, considering vehicular fleet composition in the studied area. The study established a service offering of predictive maps of atmospheric pollutant dispersion, leveraging the Graz Lagrangian Model (GRAL) and accounting for meteorological conditions and city morphology. The experimental results demonstrated accurate modeling of traffic flows; however, the prediction of air pollutants exhibited a general underestimation, attributed to input data limitations.

The work in [32] introduced a methodology for analyzing pollution emissions in a medium-sized city, focusing on minimizing exhaust emissions through modern traffic simulations. Microscopic traffic simulations were performed using the SUMO tool, enabling the accurate identification of traffic organization changes in pollution emissions before implementation. That approach ensures a smooth vehicle flow and reduced exhaust emissions. Experiments, coupled with visual modeling of traffic for pollution emissions, were executed on a key city artery in Czestochowa, Poland. The obtained results were instrumental in demonstrating the benefits of planned roadworks, indicating to the city government the imperative need for communication network modernization. The presented approach differs from our proposal since it did not include a comparison with a real route.

Finally, it is noticeable that previous studies have explored promising approaches and methodologies to understand and quantify the environmental impact of the transportation sector, as well as proposed effective measures for emission reduction. In general, some works have utilized gas sensors near a vehicle's exhaust system to collect emission data, while others have relied on machine learning algorithms, such as neural networks, to predict emissions based on real-time data from vehicle systems. These works have also highlighted existing gaps in this field and the need for novel solutions. In this context, the current article distinguishes itself by proposing a methodology to estimate CO₂ emissions using an artificial intelligence module focused on TinyML. Moreover, a real-world case study was conducted to compare emissions between gasoline and ethanol. This approach fills gaps in the literature and promotes the development of sustainable solutions for vehicle emission monitoring.

3. Proposed Approach

In this section, the practical and integrated implementation of our innovative approach to analyzing vehicle emissions in real-world and simulated environments is presented.

3.1. Real-World Monitoring

The proposed methodology in this article aims to estimate the amount of CO₂ emitted during a specific route through data collected from a target vehicle. A total of 153,255 were collected from the real scenario. This real-world element of the proposed approach involved the instrumentation between On-Board Diagnostics (OBD-II) and a smartphone to gather the necessary vehicle data, as well as centralized processing that can be performed on dedicated servers or via cloud-based services. The process flow is detailed in Figure 1.

After data collection, two processing modules were defined to estimate the CO₂ emissions.

- **Module 1—Estimating CO₂:** This module is responsible for calculating continuous CO₂ emissions based on sensor variables, notably the manifold absolute pressure (MAP) and the mass airflow (MAF). It is important to note that specific vehicle models may have different available sensors: while some vehicles are equipped with only an MAP sensor, some have only an MAF sensor, and some models have both. To handle these variations, when a vehicle lacks an MAF sensor, the estimation of CO₂ emissions is carried out using an MAP sensor to estimate the MAF [19].

- **Module 2—Data Analysis and AI Application:** After estimating the CO₂ emissions in Module 1, this cloud-based module utilizes data analysis and AI algorithms to examine the emissions patterns in relation to the type of employed fuel in the analyzed vehicle.

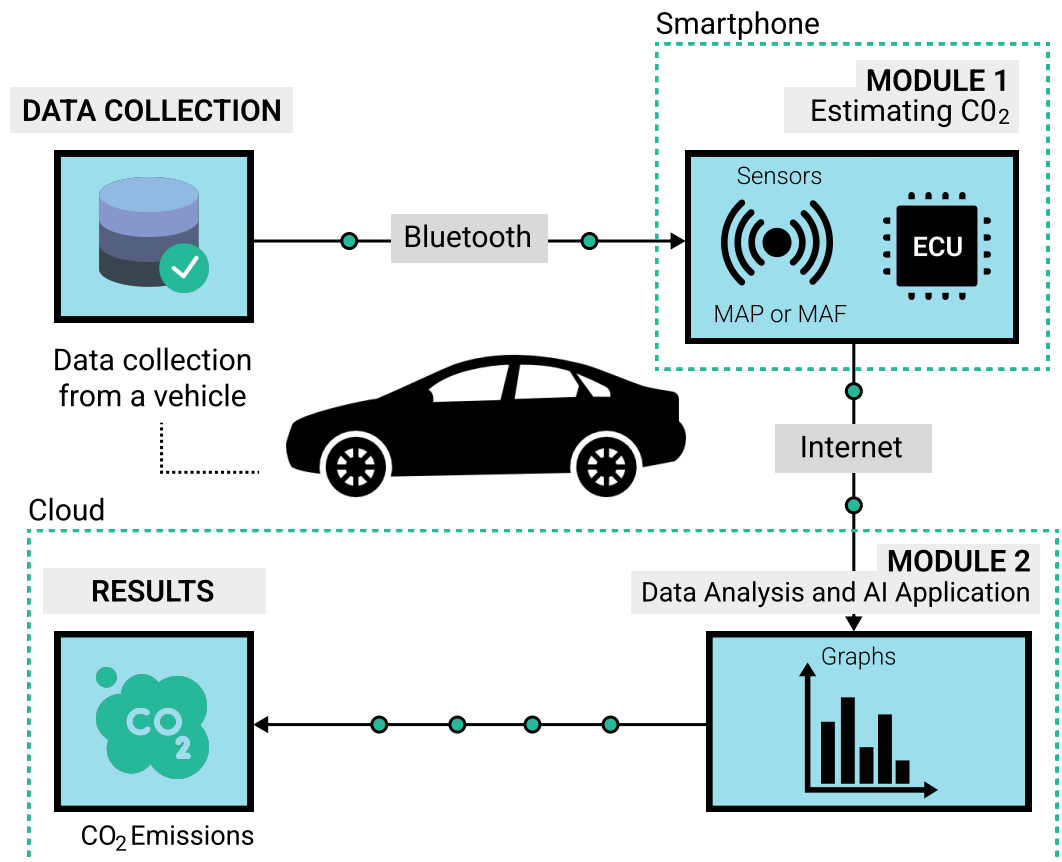


Figure 1. Overview of the proposed data processing approach for real-world monitoring.

3.1.1. Estimating CO₂

In this article, the estimation of CO₂ is performed through direct access to data using the mass airflow (MAF) sensor as a reference. With this data, the amount of fuel mass injected into the combustion chamber (C_{comb}) is calculated using Equation (1):

$$C_{comb}[\text{g/s}] = \frac{m_{af}[\text{g/s}]}{\text{AFR}} \quad (1)$$

where m_{af} represents the MAF, and the air–fuel ratio (AFR) is determined using data collected from the OBD system. Based on these variables, some conversions are performed, as expressed in Table 1, according to previous analyses [30,33]. In addition to AFR, other relevant fuel data include its density (ρ_{comb}) and the amount of CO₂ generated after burning 1 L of fuel (CO_{2PL}).

Table 1. Conversion constants.

Fuel	(ρ_{comb})	CO _{2PL}
Gasoline	737 g/L	2310 g/L
Ethanol	789 g/L	1510 g/L

In the next step, the fuel volume (V_{comb}) can be determined using Equation (2):

$$V_{comb}[\text{L/s}] = \frac{C_{comb}[\frac{\text{g}}{\text{s}}]}{\rho_{comb}[\frac{\text{g}}{\text{L}}]} \quad (2)$$

Once we have the fuel flow rate, we can finally estimate the CO₂ emissions per second using Equation (3) by multiplying V_{comb} by the CO_{2PL} coefficient.

$$\text{CO}_2[\text{g/s}] = V_{comb}[\text{L/s}] \times \text{CO}_{2PL}[\text{g/L}] \quad (3)$$

3.1.2. AI-Based Data Analysis

From the obtained estimation, a comparative evaluation of the CO₂ emissions generated through the use of gasoline and ethanol could be performed. Initially, the evaluation was carried out by applying the TEDA (Typicality and Eccentricity Data Analysis) algorithm, which is used to detect outliers in data sets [34]. This algorithm is based on the notions of typicality and eccentricity in order to increase the relevance of the obtained results.

Considering an input $x_k \in \mathbb{R}$ at a discrete time instant k , eccentricity ($\zeta_k(x_k)$) measures the difference of a sample with respect to the rest of the set, while typicality ($\tau_k(x_k)$) measures the similarity of a sample with the rest of the set. Both eccentricity and typicality can be rewritten, allowing the calculations to be performed recursively.

As these measures express opposite ideas, one can be written as the complement of the other, as expressed in the following equations.

$$\zeta_k(x_k) = \frac{1}{k} + \frac{(\mu_k - x_k)^T(\mu_k - x_k)}{k\sigma_k^2}, \quad k > 2 \quad (4)$$

$$\tau_k(x_k) = 1 - \zeta_k(x_k) \quad (5)$$

$$\mu_k(x_k) = \frac{k-1}{k}\mu_{k-1} + \frac{1}{k}x_k, \quad \mu_1 = x_k \quad (6)$$

$$\sigma_k^2(x_k) = \frac{k-1}{k}\sigma_{k-1}^2 + \frac{1}{k-1}|x_k - \mu_k|^2, \quad \sigma_1^2 = 0 \quad (7)$$

where $\mu_k(x)$ represents the mean, and σ_k^2 represents the variance for instant k . Then, both eccentricity and typicality can be normalized, as shown in Equations (8) and (9).

$$\zeta_k(x_k) = \frac{\tilde{\zeta}_k(x_k)}{2}, \quad \sum_{i=1}^k \zeta_i(x_k) = 1, \quad k \geq 2 \quad (8)$$

$$t_k(x_k) = \frac{\tau_k(x_k)}{k-2}, \quad \sum_{i=1}^k t_i(x_k) = 1, \quad k \geq 2 \quad (9)$$

Finally, an approach to identifying an outlier for any data distribution is Chebyshev's inequality, described in Equation (10).

$$\zeta_k(x_k) \geq \frac{m^2 + 1}{2k} \quad (10)$$

In this expression, m is the number of standard deviations from the mean μ_k , and it can be understood as the detection sensitivity threshold. If the aforementioned condition is true, the sample is considered an outlier, and thus, it can be ignored when computing CO₂ estimations, making the results as a whole more accurate.

At this point, vehicular pollution estimations based on the actual processing of data retrieved from vehicles could be performed, allowing comparisons of different employed fuels.

3.2. Simulated Scenarios

Although a practical mechanism for real-world monitoring is proposed, extensive experimentation can be costly, especially when long-distance journeys are considered. Therefore, we wanted to enable the simulation of realistic scenarios without the need to invest time and resources in real experiments, complementing the achievable results. Actually, it is an efficient strategy for modeling and understanding complex variables in a controlled and virtual environment. To implement this, the use of a simulator is essential, and we chose SUMO for its capability to generate detailed simulations. Moreover, its compatibility with the Python 3.10 programming language can be highlighted, particularly through the *traci* library. Finally, SUMO's user-friendly interface and flexibility for integrations with advanced programming tools are also among its favorable factors.

In order to allow computations of CO₂ gas emissions in the simulations, it was necessary to use the two vehicular sensors, MAF and AFR (Equation (1)), but they are not available in SUMO. Therefore, the training of machine learning models using variables available in both environments—the real and the simulated one—was defined. It is noteworthy that the training data for the models came from the case study highlighted in Section 4. Thus, an intersection of the variables existing in both scenarios was applied, creating a hybrid dataset that can be used to train AI models, as can be seen in Figure 2.

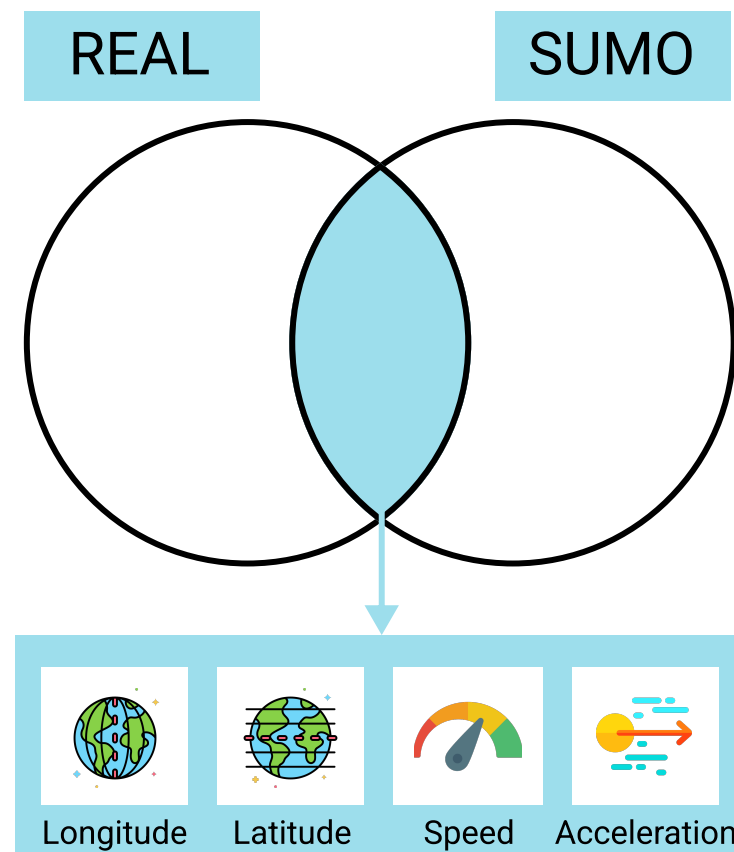


Figure 2. Intersection of existing variables in both forms of scenarios.

As a result of this process, four distinct AI models were obtained. Two of these models were designed to predict MAF and AFR values in the scenario using gasoline, while the other two models focus on predicting the same parameters but in the scenario where ethanol is the employed fuel. This approach allows for a more precise analysis tailored to the specificities of each type of fuel, providing insights into the environmental impact and efficiency of different automotive fuels.

Let us continue with the modeling process. The adopted strategy involved training four distinct models. These models were fed with a set of carefully selected variables: latitude, longitude, speed, and acceleration. These variables were chosen for their commonality in both real-world and simulated scenarios.

The training process of these models was significantly enhanced through the use of the Lazy Predictor library, an advanced tool in the field of data science [35]. This library facilitates the automation of the training process, allowing for the efficient and systematic generation and evaluation of multiple regression models. The Regressor class, a key feature of this library, was employed to build and test a variety of predictive models.

During the training phase, the mentioned library automated the training process, generating a broad range of models for each of the four key variables. After the training had been completed, the model with the best performance for each set of variables was selected. Figure 3 represents the test results for the best-performing models.

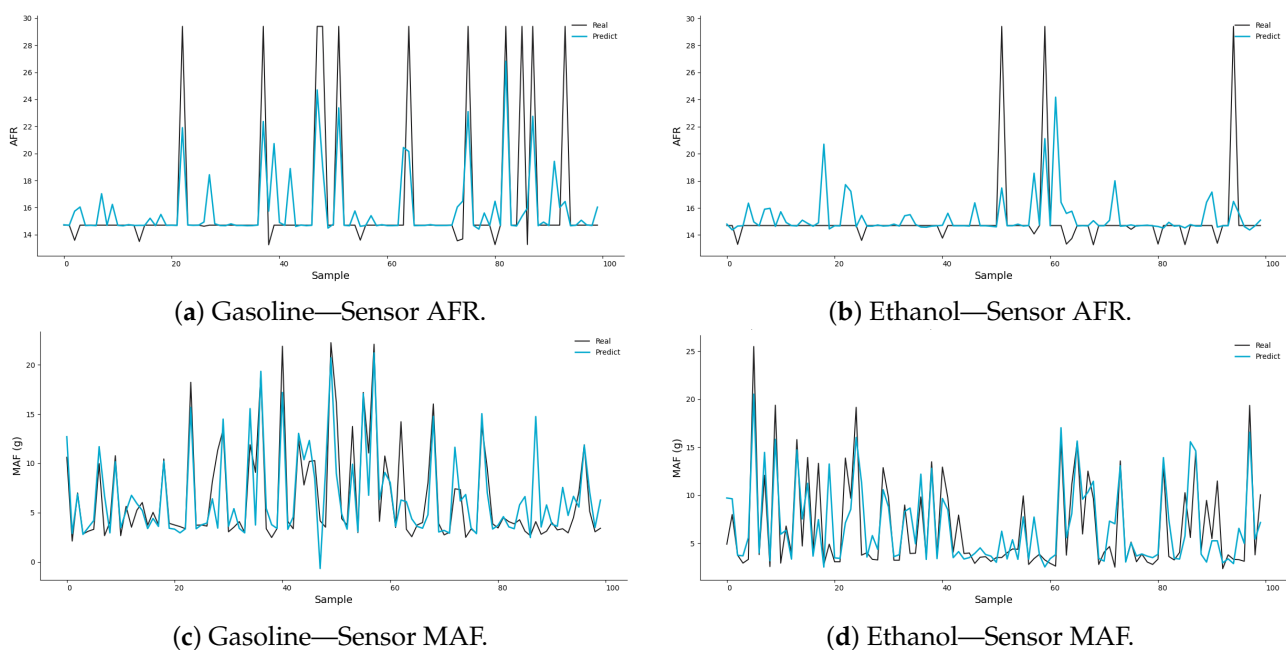


Figure 3. Test results for the best-performing models.

Finally, the two selected models for predicting MAF were of the XGBRegressor type, and those chosen for predicting AFR were of the LGBMRegressor type. This selection was based on performance metrics related to the models' errors. Figure 4 shows how processing occurs for the data that pass through each of the models, followed by the utilization of the emission calculation discussed earlier.

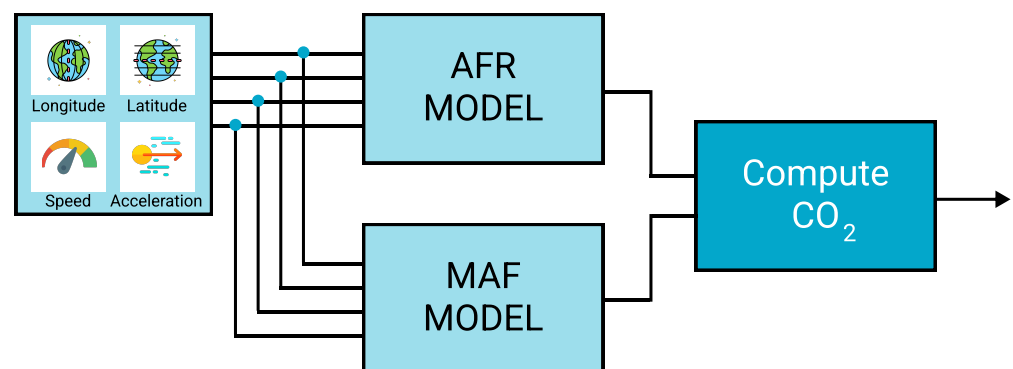


Figure 4. Data flow through XGBRegressor and LGBMRegressor models for emission calculations in simulations.

4. Case Study

In this section, the practical application of the proposed approach is explored through a case study in both real-world and simulated environments.

4.1. Experimental Scenario

A case study was considered to evaluate the proposed methodology in order to investigate the feasibility of analyzing the estimated CO₂ emissions along a route with a compressed machine learning model on different dates, using ethanol and gasoline in a flex-fuel (hybrid) vehicle. As previously mentioned, the results of this analysis can contribute to indicators for smart cities in terms of sustainability and the energy transition, specifically regarding the importance of biofuels. Since this is a real-world experiment, potentially closer to actual reality, this was the first scenario to be defined.

The following subsections describe the data collection, evaluation metrics, and execution process for this scenario.

4.1.1. Data Collection

The data collection process was conducted in a real-world scenario, with a volunteer acting as the driver of a Nissan Kicks 2022 car model with automatic transmission. The instrumentation setup was then defined, which involved configuring the environment to collect data from this vehicle. The following components were utilized:

- **OBD-II scanner:** A device was used to collect data from vehicle sensors, which were, in our case, the speed, MAP, and AFR values. The popular ELM-327 OBD-II scanner was used with a sampling rate of 1 s between each request;
- **Smartphone:** A device used for communication between OBD-II and the associated modules, as well as for storing GPS positions. The volunteer used an Android smartphone with sufficient processing, memory, and communication capabilities for the experiments;
- **Torque Pro App:** A mobile application used to facilitate the communication of the data collected via OBD-II and cloud-based applications.

Before the volunteer began the defined route, an OBD-II reader was connected to the vehicle and paired with the driver's mobile device via Bluetooth communication. Additionally, the Torque Pro App was configured to collect speed and MAF data, which were available for the vehicle in use. During the trip, the Torque Pro App 1.12.101 recorded data into a CSV file, which was transmitted to a cloud server at the end of the route for further analysis.

For the data collection procedure, a route of approximately 13 km was selected in the city of Natal, Brazil. The route encompassed urban areas with paved and asphalted sections and was conducted from 6:00 to 7:00 in the morning. The route was executed under two scenarios: one with the vehicle running on gasoline and another with ethanol. Each type of fuel was tested on five different days of the week (from Monday to Friday), resulting in a total of ten trips (five for each fuel type). Finally, after completion, all the stored data could be transmitted and processed to generate graphs using geolocation metadata.

4.1.2. Data Analysis

After applying the proposed approach to calculating CO₂ emissions, the TEDA algorithm was used to analyze the instantaneous values related to the amount of gas produced by the vehicle. In this context, the presence of outliers in each fuel type was investigated. It is important to highlight the influence of the parameter m in Chebyshev's inequality for anomaly detection. Therefore, understanding the relationship between the parameter m and anomaly detection is crucial for interpreting the results.

The parameter m acts as a sensitivity threshold, setting the allowable range for values that are considered outliers. Its influence is visualized graphically in Figure 5.

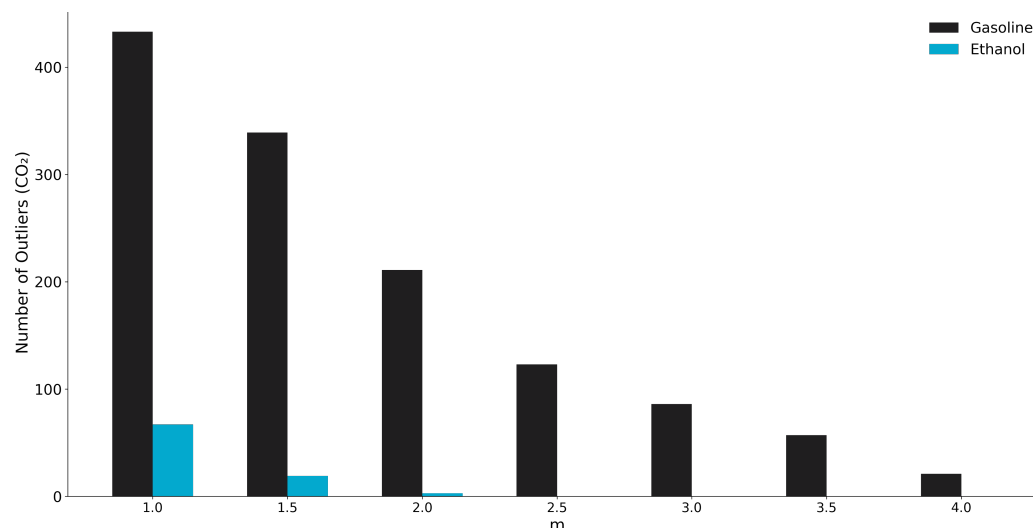


Figure 5. CO₂(g) outliers detected based on the value of parameter m .

Figure 5 graphically illustrates this influence, demonstrating that an increase in m leads to less sensitivity to extreme values, while a decrease in m increases the sensitivity to the presence of outliers. This principle then guides the selection of outliers for exclusion, making the results potentially more meaningful.

4.2. Simulated Scenario

The simulated scenario aimed to replicate the real-world data collection procedure using SUMO as the traffic simulator, enhancing the achieved results for better analysis. In the virtual environment, a scenario that mimics the urban layout and traffic conditions of the chosen route in Natal, Brazil, was configured, adopting the following configurations:

- **SUMO configuration:** The simulation was configured to replicate the urban route with details such as the road layout, intersections, and traffic density. The vehicle type was specified as a flex-fuel hybrid model;
- **OBD-II equivalents:** Virtual OBD-II equivalents were created in SUMO to mimic the data collection from the vehicle sensors. The speed, MAP, and AFR parameters were simulated with patterns resembling those expected in a real-world scenario;
- **Geolocation data:** A graph representing the geographic behavior of the city of Natal, Brazil, was created. Such a graph is essential to ensure that the simulator accurately reflects the real conditions of the city's urban roads. The creation of this graph began with the use of the Python library OpenStreetMap nx (OSMnx), a tool for manipulating and analyzing geographic data. With OSMnx, it was possible to extract a detailed map of the streets, avenues, and other relevant geographic features of Natal.

Therefore, it is worth highlighting that, for the route simulation, the result was a comprehensive graph that captured the complexity and specificity of the city's road network, as illustrated in Figure 6.

However, to guarantee the compatibility of the graph with SUMO, an additional conversion step was necessary. To do this, the `netconvert` tool was included in the SUMO installation package. This tool was designed to transform graphs of different formats into a layout that is compatible with SUMO, facilitating the integration between the simulation environment and the real geographic data.



Figure 6. Map capturing the specificity of the city's road network during the simulation.

4.3. Evaluation Metrics

The evaluation of the proposed approach required the use of specific metrics to assess the expected outcomes. The employed metrics for this evaluation were the mean absolute error (MAE) and the root mean squared error (RMSE), which both provide insights into the precision of the predictive models in capturing the variations in CO₂ emissions along a simulated route.

The adopted evaluation metrics are expressed as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^N |x_i - \hat{x}_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^N [(x_i - \hat{x}_i)]^2} \quad (12)$$

The simulated scenario was executed for both gasoline and ethanol fuels, with multiple runs to capture variations. The goal was to ensure that the simulated data reflected the diversity observed in the real-world scenario, allowing valuable comparisons. In this way, a total of 112,964 records relating to gasoline consumption and 40,291 records relating to ethanol consumption were collected from the real scenario. The predominance of gasoline use data indicates a greater representation of this fuel in the sample. To build a model, the collected data were separated, with 80% intended for training and 20% for testing, providing an adequate division to evaluate the effectiveness of the model in both situations.

The data generated from the simulation were then saved in a format similar to the one applied to the real-world scenario (CSV), allowing for a comparative analysis of CO₂ emissions and other relevant parameters. This process provided a comprehensive evaluation of the proposed methodology under controlled and repeatable conditions.

The proposed methodology was made readily accessible for research and practical purposes. The detailed implementation of our method is publicly available on our GitHub repository. This open-access approach is intended to facilitate collaboration, replication, and further research endeavors within the academic and professional communities. To access the full implementation, please visit our GitHub repository at <https://github.com/conect2ai/MDPI2023-pollution> (accessed on 12 January 2024).

5. Results

This section aims to provide a detailed description of the results obtained in both the real-world and simulated scenarios, offering a comprehensive overview of the outcomes of the defined case.

5.1. Practical Experimentation

First, in order to conduct a more accurate comparative analysis, the initially collected 2000 data samples from each of the 10 created datasets (5 for ethanol and 5 for gasoline, assuming that each day of the experiment was processed separately) were selected to ensure equivalence in the amount of processed data.

Through this methodologically established approach, the goal was to gain a deep understanding of the effects resulting from the choice between ethanol and gasoline, taking into consideration their direct influence on CO₂ emissions.

Initially, to examine the behavior of outliers in each type of fuel, the TEDA algorithm was applied with the value of $m = 1.5$. The achieved results can be observed in Figure 7.

According to Figure 7, it can be observed that there was a higher number of outliers in the gasoline data. This finding can be interpreted as an indication that the use of gasoline may result in a more heterogeneous CO₂ emission pattern, exhibiting a greater dispersion around the mean values.

For a more in-depth investigation and to corroborate this statement, it is pertinent to use a distribution plot to examine the distribution of CO₂ emission values for each type of fuel. The visualizations presented in Figure 8 depict the kernel density estimation (KDE) curve, a statistical technique that estimates the density of a variable through smoothing, generating a continuous estimate.

Upon analyzing the results in these figures, the heterogeneity of emission values related to gasoline becomes evident, as indicated by the flatter curve. As previously mentioned, this suggests that the CO₂ emission values associated with gasoline exhibit a greater dispersion around the mean. In the case of ethanol, which has fewer outliers, the KDE curve tended to concentrate more around the mean, indicating lower variability in CO₂ emission values.

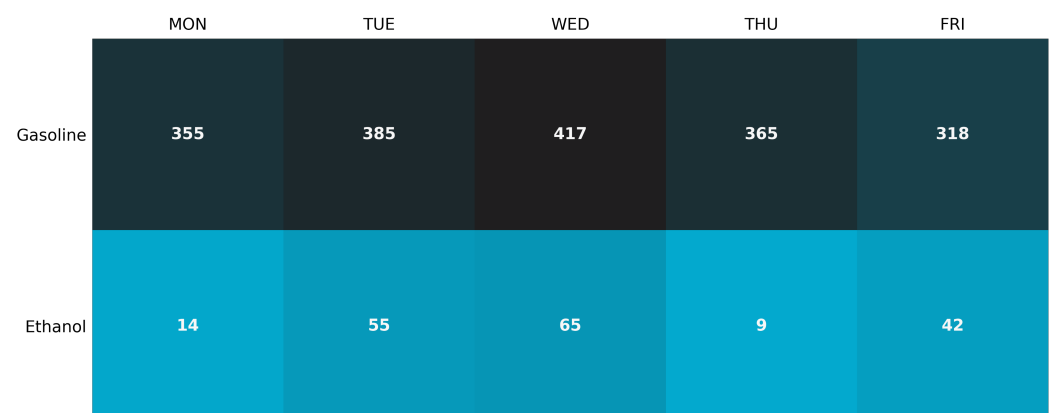


Figure 7. Sample outlier detection via TEDA.

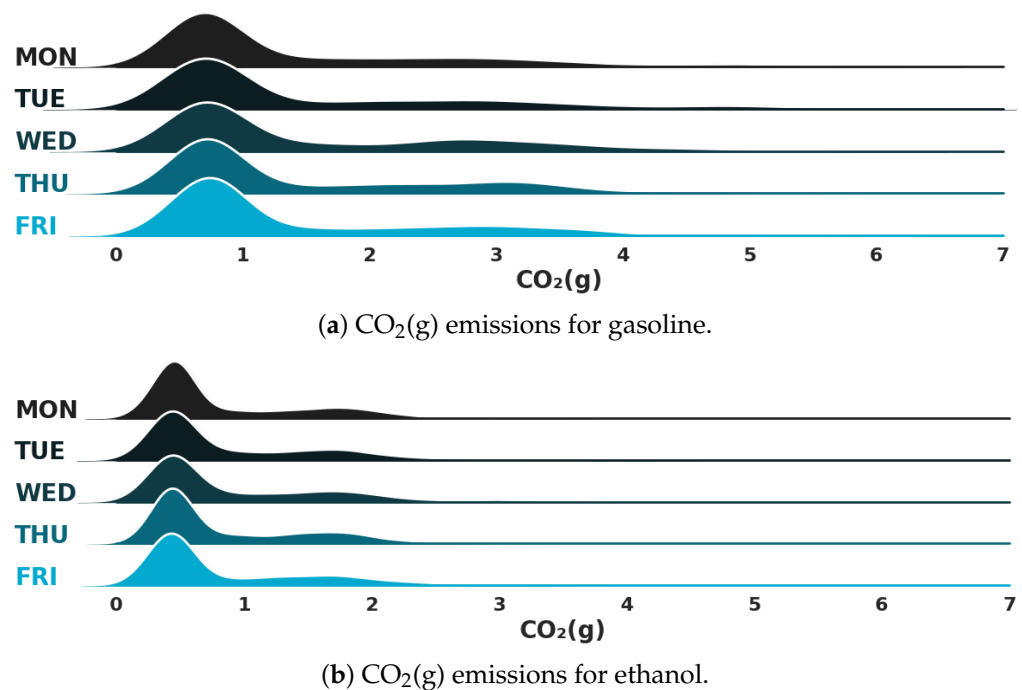


Figure 8. KDE distribution of CO₂(g) emissions for gasoline and ethanol.

An additional piece of information highlighted in Figure 8 is that gasoline, on average, exhibited higher CO₂ emissions. This observation becomes clearer when examining Figure 9.

In Figure 9, a graphical representation of the average CO₂ emissions for each type of fuel separated by weekdays is displayed. It can be observed that the average for gasoline was at a higher level compared to the average for ethanol. This indicates that, in general, the use of gasoline resulted in higher average CO₂ emissions than the use of ethanol.

While the mean was heavily influenced by the presence of outliers, Figure 10 provides evidence that there will indeed be a significantly higher CO₂ emission from gasoline throughout the performed trip.

Figure 10 was generated from the average of the first 2000 data samples for each day, corresponding to each type of fuel. This graphical representation highlights how, over time, the cumulative emission of gasoline was substantially higher than that of ethanol.

When observing Figure 10, it can be noticed that the curve corresponding to the cumulative emission of gasoline had a more pronounced upward trend compared to the ethanol curve. This indicates that, on average, the CO₂ emission associated with the use of gasoline accumulated in larger quantities over the analyzed period compared to ethanol, which reinforced the urgency of reducing its use as a fuel in combustion-engine vehicles [26].

5.2. Simulated Experiments

The results obtained from the simulations demonstrated remarkable conformity with the data collected from real scenarios, highlighting the effectiveness of the simulated environment in replicating authentic driving conditions, as can be seen in Figure 11.

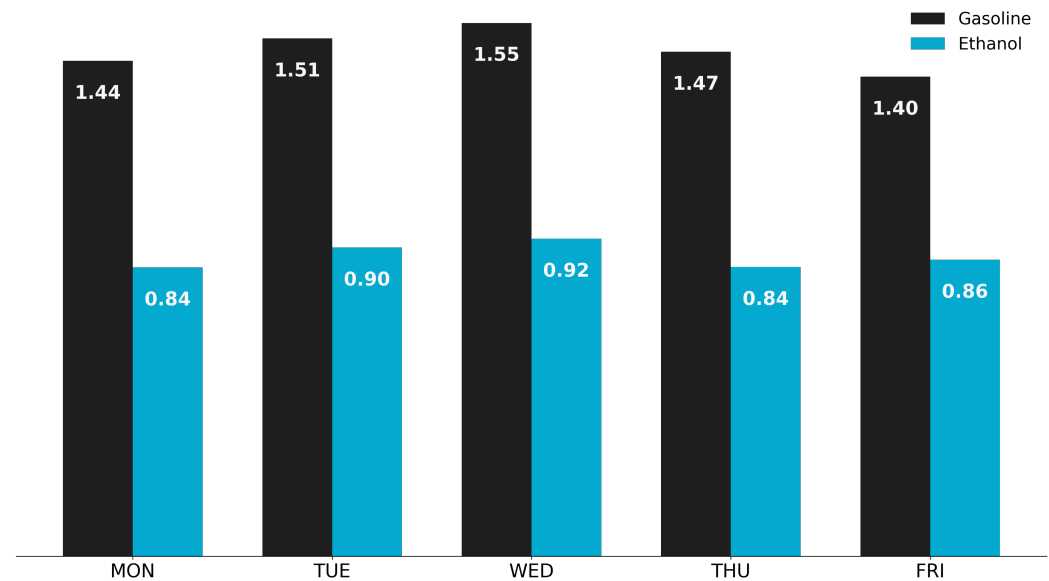


Figure 9. Average CO₂(g) emissions by weekday.

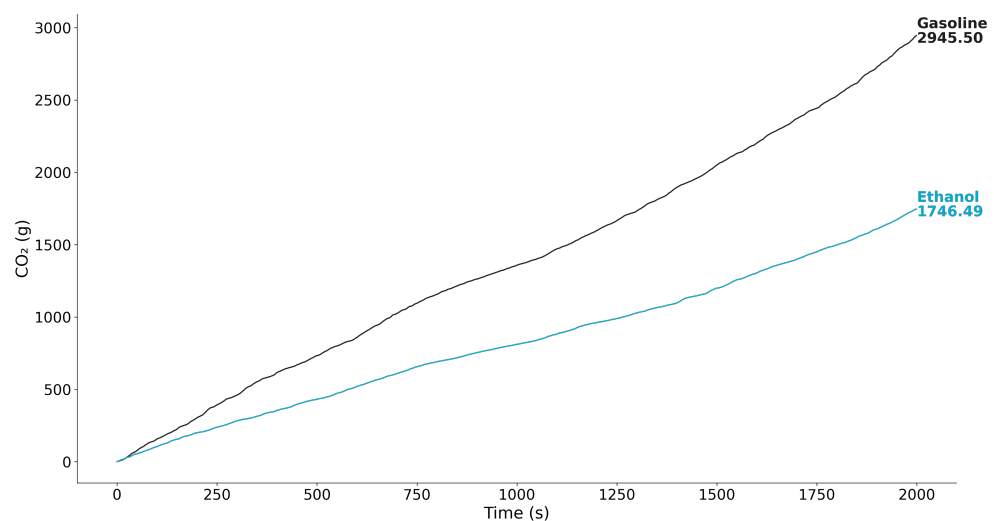


Figure 10. Average cumulative sum of CO₂(g) emissions.

In this way, Figure 11 indicates the achieved results for different driving scenarios. For the simulation, some points where the vehicle simulated in SUMO should cross were manually selected, which were also points crossed by the real vehicle. SUMO uses graph optimization techniques to search for the shortest distance between each of the two selected points. In other words, these points were selected in such a way that they replicated the actually selected route with the difference that the one selected for this stage was shorter (but with no practical impact on the performed analysis).

Furthermore, a comparison of the emissions generated through the simulated environment using the developed modules was carried out, as can be seen in Figure 12, which presents a comparative analysis of the accumulated sum of gasoline and ethanol emissions. Consistent with the initial graphical representation, the cumulative emissions from gasoline use were higher than those from ethanol. This difference in cumulative emissions is represented visually in the graph, which delineates the disparity between the two fuel types.

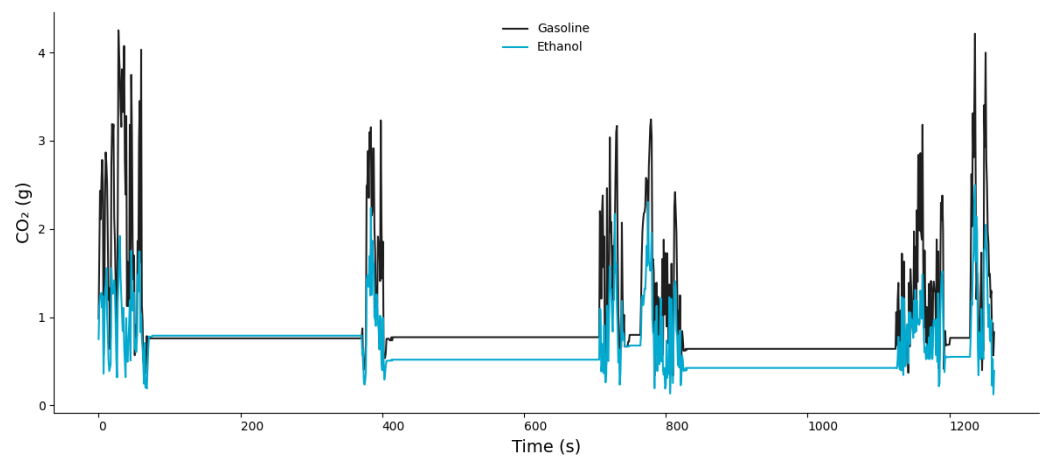


Figure 11. Comparison of emissions for gasoline and ethanol, showcasing the remarkable conformity between the considered evaluation scenarios.

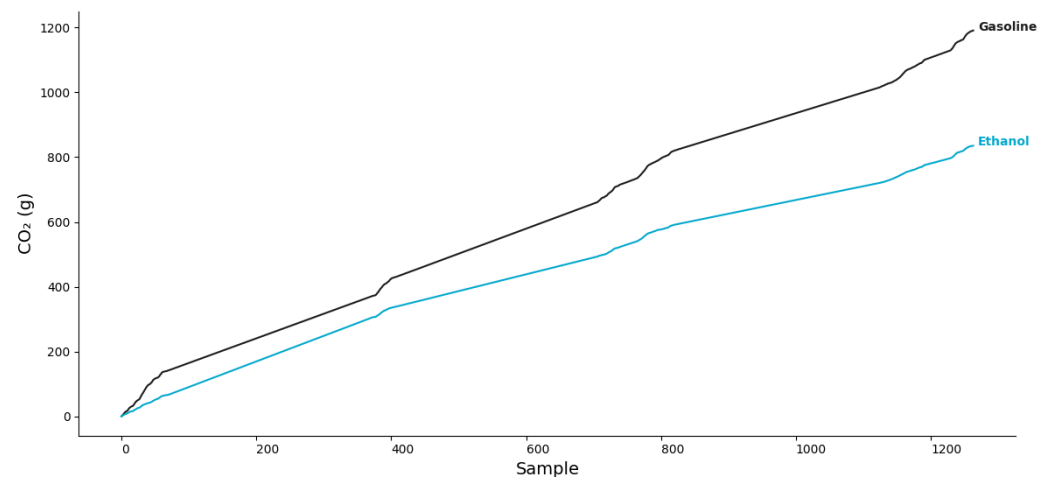


Figure 12. Comparison of emissions for gasoline and ethanol, showcasing the remarkable conformity between simulation results.

Even though Figures 10 and 12 do not depict the same route, a resemblance can be observed in the generated graphs, indicating a similarity in the behavior replicated by the simulated environment. An effective way to compare the impact of fuels is through map visualization, as exemplified in Figure 13. In these visualizations, the complete emission data for a single day (Monday) were considered for both real-world and simulated scenarios. In this case, the employed simulator played an integral role in representing real-world conditions. SUMO, in particular, stood out for its ability to incorporate a comprehensive range of geographic and structural road characteristics. These elements, when combined with AI models, allowed the creation of an extremely realistic simulated environment. The simulations were able to capture the complexity of the interactions between the vehicle, the driver, and the environment, thus providing a tool for analyzing CO₂ emissions.

In Figure 13, it is noticeable that both graphs show reddish shades in similar regions, which is an indication of higher emitted CO₂ in those areas. This observation can be attributed to the fact that the car dynamics tended to behave similarly in both cases. This reinforces the idea that a simulated environment can be a viable approach to generating additional data that resemble similar characteristics. However, it is still evident that the shades for gasoline tended to be much closer to the colors indicating higher CO₂ emissions.

Gasoline

Real

Simulation

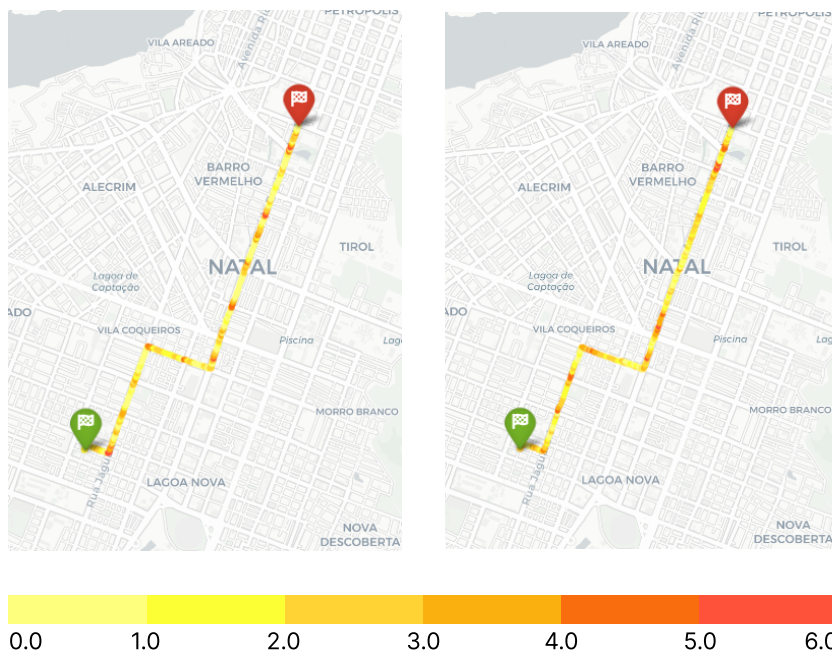


(a) Map view of emissions for gasoline.

Ethanol

Real

Simulation



(b) Map view of emissions for ethanol.

Figure 13. Comparison of emission maps for real-world and simulated scenarios.

Further analysis concerned a comparison using the calculation previously presented but also incorporating data from the AFR and MAF sensors. Figure 14 provides a visual analysis of the AFR and MAF metrics, highlighting how they responded to the adopted calculation method. This close alignment between the simulated data and the real data reinforces the feasibility of using simulators and AI for advanced studies in the field of automotive and environmental engineering.

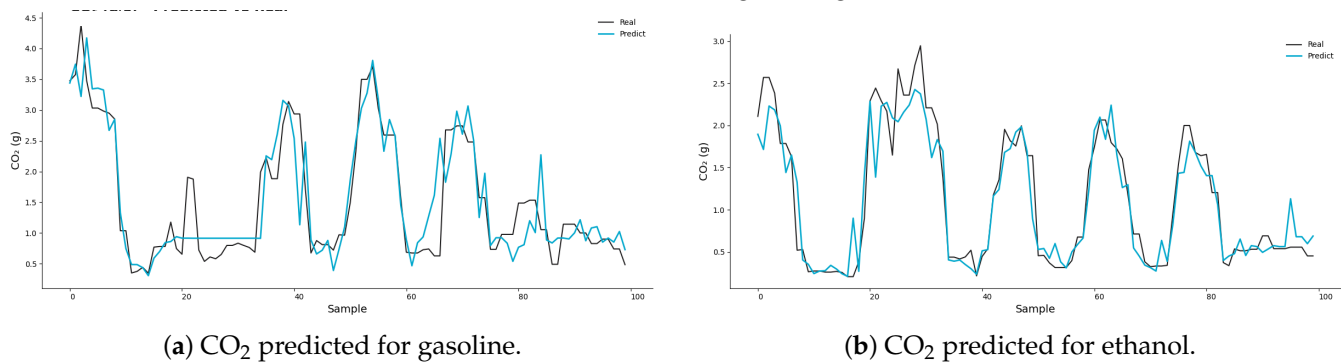


Figure 14. Comparison of CO₂ predictions in the real world.

Finally, for the conducted simulation study comparing the emissions of two different fuel types, important results are presented in Table 2.

Table 2. MAE and RMSE for gasoline and ethanol.

Fuel Type	MAE	RMSE
Gasoline	0.4151	0.6222
Ethanol	0.2334	0.3624

First, considering the defined evaluation metrics, the simulated (predictive) model for ethanol exhibited a substantially lower MAE (0.2334) compared to that for gasoline (0.4151). This indicates that, on average, the predictions for ethanol emissions were closer to the actual values, signifying a higher level of accuracy in replicating real-world conditions in the simulation. Further emphasizing the model's performance, the RMSE values reinforce the superiority of the simulated ethanol model. With RMSEs of 0.3624 for ethanol and 0.6222 for gasoline, the smaller RMSE for ethanol signifies a more precise representation of CO₂ emission variations in the simulated scenario. Therefore, it emphasizes the potential of the proposed methodology to assess the environmental impacts of different fuel types in a simulated urban environment with the use of digital twins performing satisfactorily well in the defined scenario.

5.3. Discussions and Analyses

The results of both the real-world and simulated experiments provide a nuanced understanding of the implications associated with the choice between ethanol and gasoline in terms of CO₂ emissions.

The outlier analysis revealed a higher number of outliers in gasoline emissions, suggesting a more heterogeneous emission pattern. This variability could have significant implications for environmental planning and policy-making, as it indicates that gasoline-powered vehicles may contribute to a less consistent level of CO₂ emissions compared to ethanol.

The consistently higher average and cumulative emissions for gasoline underscore its greater impact on the environment. This aligns with existing knowledge about the carbon footprint of gasoline and emphasizes the urgency of transitioning to more sustainable fuel alternatives.

The accuracy of the simulation environment and the superior performance of the predictive model for ethanol suggest that ethanol might be a more environmentally friendly alternative, at least in terms of CO₂ emissions. This conformity between simulated and real-world data is crucial for predicting and understanding the environmental impact of different fuels.

Considering these patterns, there are important implications for environmental policies and initiatives. Policymakers might need to prioritize promoting the use of ethanol or other alternative fuels to reduce the overall carbon footprint. Additionally, this study's findings might encourage behavioral changes, such as a shift towards cleaner energy sources or more sustainable transportation practices.

Thus, this study highlights the need for careful consideration when choosing between ethanol and gasoline. The environmental consequences, as evidenced by higher emissions from gasoline, should play an important role in decision-making processes. By informing policy-makers, encouraging behavioral changes, and guiding future research directions, this study contributes to a more comprehensive understanding of the environmental implications of fuel choices.

5.4. Research Limitations and Challenges

In this section, we discuss some of the limitations identified in our research.

(a) Sample Size and Study Duration:

- Sample Size: The initial sample of 2000 data points per day may be deemed limited in capturing the full diversity of driving conditions;
- Study Duration: While the analysis period was sufficient for the study's objectives, it may not have encompassed seasonal variations or long-term effects that could influence emissions.

(b) Simulation Limitations:

- Model Complexity: The complexity of the simulation model may not fully reflect the intricacies of real driver and traffic behavior, potentially impacting simulated emissions;

(c) Geographical Representation and Fuel Variations:

- Geographical Representation: Despite the simulation incorporating geographical features, the complete representation of topography and road infrastructure may not be entirely accurate;
- Fuel Composition Variations: Variations in ethanol and gasoline composition may not have been fully addressed, and different fuel blends may have resulted in distinct emissions.

(d) Unconsidered External Factors and Implicit Bias:

- Unconsidered External Factors: The study may not have fully considered external factors, such as weather conditions, that can influence emissions and were not controlled for;
- Implicit Bias in Modeling: The modeling may reflect certain driving behaviors or decisions influenced by implicit biases present in the original dataset.

6. Conclusions

This article has presented an IoT-based approach that employed a smartphone, a mathematical model, and an AI algorithm to estimate CO₂ emissions during vehicle operation, conducting intelligent analysis of the results. In addition, we employed SUMO to create a simulation scenario powered by a linear regression AI model trained with data collected via the IoT approach, which faithfully reflected the real operating conditions of the vehicles and enhanced the set of achieved experimental results. Thus, it was possible to evaluate the effectiveness of two different types of fuels, making it easier to understand the environmental implications arising from the choice of different fuels in the automotive sector.

A case study compared the emissions of ethanol and gasoline fuels, highlighting that ethanol exhibits significantly lower CO₂ emissions, emphasizing the importance of more sustainable fuels in reducing environmental impacts and mitigating climate change. In the simulated environment, SUMO's detailed configuration, including flex-fuel modeling and the creation of OBD-II virtual equivalents, enabled controlled and repeatable analysis. The efficient conversion of real geographic data to the SUMO-compatible format was essential to ensure simulation fidelity. The final outcome was a comprehensive analysis of air pollution due to combustion engine vehicles, which may be highly significant when fostering the transition to more sustainable transportation.

As an additional result, the inclusion of evaluation metrics such as the mean absolute error (MAE) and root mean squared error (RMSE) significantly enriched our analysis, offering quantitative insights into the accuracy of predictive models and enabling a direct comparison between the gasoline and ethanol scenarios. The attainment of low MAE and RMSE values indicates that, on average, our models yielded predictions in close proximity to the actual CO₂ emission values, underscoring a high degree of accuracy in replicating emission variations. This numerical precision is particularly crucial when discerning between the two fuel types, with the ethanol scenario exhibiting notably lower errors compared to gasoline. These metrics not only enhance the robustness of our findings but also provide a concise and quantitative measure of the reliability of our predictive models, contributing valuable information for informed decision-making and policy formulation in the context of mitigating CO₂ emissions. These metrics are important in understanding how predictions align with actual variations in CO₂ emissions along a route. Additionally, it is crucial to address a specific limitation related to the MAF sensor, which served as a reference for CO₂ emissions estimation in our approach. As highlighted, since our methodology relies on sensor data, we acknowledge the potential impact of sensor failures on estimation accuracy. Therefore, maintaining the proper functioning of sensors is paramount to ensure the reliability of our methodology.

Future works will incorporate this proposed approach into OBD-II Edge devices as a TinyML solution, which would operate autonomously and eliminate the need for smartphones, enabling more practical implementation. This could even allow more widespread dissemination of air pollution monitoring mechanisms within a smart city ecosystem, with adaptive urban services responding to increased pollution levels by diverting traffic or imposing temporary limitations for combustion-engine vehicles. Additionally, it is essential to expand the possible set of analyses to different types of vehicles, considering their specificities in terms of CO₂ emissions, and increase the sample size in the number of both vehicles and routes to achieve a more representative understanding of vehicle emissions in diverse contexts. In this sense, the development of more generic models that can be applied to a variety of urban contexts is also intended, considering different traffic profiles and road infrastructure.

Furthermore, concerning promising future works, since the simulated model assumes a simplified representation of a vehicle, some automotive characteristics that can influence CO₂ calculations may be identified more accurately using other simulators, such as agent-based modeling (ABM). In addition, an important focus that should be applied is the analysis of potential limitations and challenges associated with the widespread adoption of ethanol as a fuel source. Issues such as refueling infrastructure, production sustainability, public acceptance, and socioeconomic impacts deserve detailed attention. By exploring these aspects, future research can contribute to a more comprehensive understanding, considering not only environmental implications but also the practical and ethical challenges related to the transition to ethanol as a more sustainable fuel alternative.

Author Contributions: Formal analysis, M.A. (Matheus Andrade), M.M., T.M. and M.A. (Mariana Azevedo); project administration, I.S.; software, M.A. (Matheus Andrade), T.M., M.M. and M.A. (Mariana Azevedo); writing, original draft, M.S. and M.A. (Matheus Andrade); research supervisors, I.S., M.S. and D.G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financed by the Brazilian fostering agency CNPq (National Council for Scientific and Technological Development), Process No. 405531/2022-2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available at the following link: <https://github.com/conect2ai/MDPI2023-pollution> (accessed on 12 January 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ABM	Agent-based modeling
AI	Artificial intelligence
AFR	Air–fuel ratio
CO	Carbon monoxide
CO ₂	Carbon dioxide
GHGs	Greenhouse gases
GRAL	Graz Lagrangian Model
IAT	Intake absolute temperature
IoT	Internet of Things
KDE	Kernel density estimation
LSTM	Long short-term memory
MAE	Mean absolute error
MAF	Mass airflow
MAP	Manifold absolute pressure
OBD	On-board diagnostics
OSMnx	Open Street Maps nx
RMSE	Root mean square error
RPM	Revolutions per minute
SUMO	Simulation of urban mobility
TCE	Transportation carbon emissions
TEDA	Typicality and eccentricity data analysis
TinyML	Tiny machine learning
UN	United Nations
VEIN	Vehicular emissions inventories

References

1. Santos, U.; Arbex, M.; Braga, A.; Mizutani, R.; Cançado, J.; Terra-Filho, M.; Chatkin, J. Environmental air pollution: Respiratory effects. *J. Bras. Pneumol.* **2021**, *47*. [[CrossRef](#)]
2. United States Environmental Protection Agency. Sources of Greenhouse Gas Emissions. Available online: <https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions> (accessed on 10 May 2023).
3. Barman, P.; Dutta, L.; Bordoloi, S.; Kalita, A.; Buragohain, P.; Bharali, S.; Azzopardi, B. Renewable energy integration with electric vehicle technology: A review of the existing smart charging approaches. *Renew. Sustain. Energy Rev.* **2023**, *183*, 113518. [[CrossRef](#)]
4. Hoang, A.T.; Pham, V.V.; Nguyen, X.P. Integrating renewable sources into energy system for smart city as a sagacious strategy towards clean and sustainable process. *J. Clean. Prod.* **2021**, *305*, 127161. [[CrossRef](#)]
5. Liu, F.; Shafique, M.; Luo, X. Literature review on life cycle assessment of transportation alternative fuels. *Environ. Technol. Innov.* **2023**, *32*, 103343. [[CrossRef](#)]
6. Huang, Y.; Zhang, Y.; Deng, F.; Zhao, D.; Wu, R. Impacts of Built-Environment on Carbon Dioxide Emissions from Traffic: A Systematic Literature Review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 16898. [[CrossRef](#)] [[PubMed](#)]
7. Gurney, K.R.; Şiir Kılıç, S.; Seto, K.C.; Lwasa, S.; Moran, D.; Riahi, K.; Keller, M.; Rayner, P.; Luqman, M. Greenhouse gas emissions from global cities under SSP/RCP scenarios, 1990 to 2100. *Glob. Environ. Chang.* **2022**, *73*, 102478. [[CrossRef](#)]

8. Fan, J.; Meng, X.; Tian, J.; Xing, C.; Wang, C.; Wood, J. A review of transportation carbon emissions research using bibliometric analyses. *J. Traffic Transp. Eng.* **2023**, *10*, 878–899. [[CrossRef](#)]
9. Aba, M.M.; Amado, N.B.; Rodrigues, A.L.; Sauer, I.L.; Richardson, A.A.M. Energy transition pathways for the Nigerian Road Transport: Implication for energy carrier, Powertrain technology, and CO₂ emission. *Sustain. Prod. Consum.* **2023**, *38*, 55–68. [[CrossRef](#)]
10. Holechek, J.L.; Geli, H.M.; Sawalhah, M.N.; Valdez, R. A global assessment: Can renewable energy replace fossil fuels by 2050? *Sustainability* **2022**, *14*, 4792. [[CrossRef](#)]
11. dos Santos, F.S.; Andreão, W.L.; Miranda, G.A.; de Carvalho, A.N.M.; Pinto, J.A.; Pedruzzi, R.; Carvalho, V.S.B.; de Almeida Albuquerque, T.T. Vehicular air pollutant emissions in a developing economy with the widespread use of biofuels. *Urban Clim.* **2021**, *38*, 100889. [[CrossRef](#)]
12. Ogunkunle, O.; Ahmed, N.A. Overview of Biodiesel Combustion in Mitigating the Adverse Impacts of Engine Emissions on the Sustainable Human–Environment Scenario. *Sustainability* **2021**, *13*, 5465. [[CrossRef](#)]
13. Sanches, G.M.; de Oliveira Bordonal, R.; Magalhães, P.S.G.; Otto, R.; Chagas, M.F.; de Fátima Cardoso, T.; dos Santos Luciano, A.C. Towards greater sustainability of sugarcane production by precision agriculture to meet ethanol demands in south-central Brazil based on a life cycle assessment. *Biosyst. Eng.* **2023**, *229*, 57–68. [[CrossRef](#)]
14. Gauto, M.A.; Carazzolle, M.F.; Rodrigues, M.E.P.; de Abreu, R.S.; Pereira, T.C.; Pereira, G.A.G. Hybrid vigor: Why hybrids with sustainable biofuels are better than pure electric vehicles. *Energy Sustain. Dev.* **2023**, *76*, 101261. [[CrossRef](#)]
15. Hopkins, E.; Potoglou, D.; Orford, S.; Cipcigan, L. Can the equitable roll out of electric vehicle charging infrastructure be achieved? *Renew. Sustain. Energy Rev.* **2023**, *182*, 113398. [[CrossRef](#)]
16. Lou, L.; Li, Q.; Zhang, Z.; Yang, R.; He, W. An IoT-Driven Vehicle Detection Method Based on Multisource Data Fusion Technology for Smart Parking Management System. *IEEE Internet Things J.* **2020**, *7*, 11020–11029. [[CrossRef](#)]
17. Manivannan, R. Research on IoT-based hybrid electrical vehicles energy management systems using machine learning-based algorithm. *Sustain. Comput. Inform. Syst.* **2024**, *41*, 100943. [[CrossRef](#)]
18. Bezerra, C.G.; Costa, B.S.J.; Guedes, L.A.; Angelov, P.P. An evolving approach to data streams clustering based on typicality and eccentricity data analytics. *Inf. Sci.* **2020**, *518*, 13–28. [[CrossRef](#)]
19. Andrade, P.; Silva, I.; Silva, M.; Flores, T.; Cassiano, J.; Costa, D.G. A tinyml soft-sensor approach for low-cost detection and monitoring of vehicular emissions. *Sensors* **2022**, *22*, 3838. [[CrossRef](#)]
20. Banbury, C.; Zhou, C.; Fedorov, I.; Matas, R.; Thakker, U.; Gope, D.; Janapa Reddi, V.; Mattina, M.; Whatmough, P. Micronets: Neural network architectures for deploying tinyml applications on commodity microcontrollers. *Proc. Mach. Learn. Syst.* **2021**, *3*, 517–532.
21. de Prado, M.; Rusci, M.; Capotondi, A.; Donze, R.; Benini, L.; Pazos, N. Robustifying the deployment of tinyml models for autonomous mini-vehicles. *Sensors* **2021**, *21*, 1339. [[CrossRef](#)]
22. Amini, S.; Orlich, C.; Beil, C.; Keler, A.; Bogenberger, K. Integrating SUMO in an urban digital twin—A case study from Munich. In Proceedings of the SUMO User Conference 2023, Berlin, Germany, 2–4 May 2023.
23. Kušić, K.; Schumann, R.; Ivanjko, E. A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics. *Adv. Eng. Inform.* **2023**, *55*, 101858. [[CrossRef](#)]
24. Bagheri, M.; Bartin, B.; Ozbay, K. Simulation of Vehicles' Gap Acceptance Decision at Unsignalized Intersections Using SUMO. *Procedia Comput. Sci.* **2022**, *201*, 321–329. [[CrossRef](#)]
25. United Nations. The Sustainable Development Goals Report. 2022. Available online: <https://unstats.un.org/sdgs/report/2022> (accessed on 8 October 2022).
26. Sandaka, B.P.; Kumar, J. Alternative vehicular fuels for environmental decarbonization: A critical review of challenges in using electricity, hydrogen, and biofuels as a sustainable vehicular fuel. *Chem. Eng. J. Adv.* **2023**, *14*, 100442. [[CrossRef](#)]
27. Akhila, R.; Amoghavarsha, B.; Karthik, B.; Prajwal, Y.; et al. Internet of Things based Detection and Analysis of Harmful Vehicular Emissions. In Proceedings of the 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 20–22 January 2022; pp. 630–636.
28. Singh, M.; Dubey, R. Deep Learning Model Based CO₂ Emissions Prediction using Vehicle Telematics Sensors Data. *IEEE Trans. Intell. Veh.* **2021**, *8*, 768–777. [[CrossRef](#)]
29. Sahay, S.; Pawar, P. An Optimal Approach to Vehicular CO₂ Emissions Prediction using Deep Learning. In Proceedings of the 2023 International Conference on Emerging Smart Computing and Informatics (ESCI), Pune, India, 1–3 March 2023; pp. 1–5.
30. Flores, T.; Silva, M.; Andrade, P.; Silva, J.; Silva, I.; Sisinni, E.; Ferrari, P.; Rinaldi, S. A TinyML soft-sensor for the internet of intelligent vehicles. In Proceedings of the 2022 IEEE International Workshop on Metrology for Automotive (MetroAutomotive), Modena, Italy, 4–6 July 2022; pp. 18–23.
31. Gonçalves, F.; Silva, G.O.; Santos, A.; Rocha, A.M.A.; Peixoto, H.; Durães, D.; Machado, J. Urban Traffic Simulation Using Mobility Patterns Synthesized from Real Sensors. *Electronics* **2023**, *12*, 4971. [[CrossRef](#)]
32. Brzozowska, A.; Korczak, J.; Kalinichenko, A.; Bubel, D.; Sukiennik, K.; Sikora, D.; Stebila, J. Analysis of Pollutant Emissions on City Arteries—Aspects of Transport Management. *Energies* **2021**, *14*, 3007. [[CrossRef](#)]
33. Signoretti, G.; Silva, M.; Andrade, P.; Silva, I.; Sisinni, E.; Ferrari, P. An Evolving TinyML Compression Algorithm for IoT Environments Based on Data Eccentricity. *Sensors* **2021**, *21*, 4153. [[CrossRef](#)]

-
34. Angelov, P. Outside the Box: An Alternative Data Analytics Framework. *J. Autom. Mob. Robot. Intell. Syst.* **2014**, *8*, 29–35. [[CrossRef](#)]
 35. Pandala, S. Lazy Predict: Build Basic Models Without Much Code. 2022. Available online: <https://github.com/shankarpandala/lazypredict> (accessed on 27 December 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.