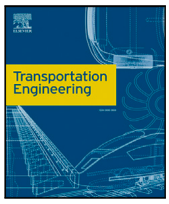
Transportation Engineering 19 (2025) 100237

Contents lists available at ScienceDirect

Transportation Engineering

journal homepage: www.sciencedirect.com/journal/transportation-engineering



Full Length Article

Promoting sustainable and personalized travel behaviors while preserving data privacy

Cláudia Brito [a,*], Noela Pina [b,c], Tânia Esteves [a], Ricardo Vitorino [b], Inês Cunha [b], João Paulo [a]

[a] INESC TEC & Universidade do Minho, R. da Universidade, Braga, 4710-057, Portugal
[b] Ubiwhere, lda., Travessa Senhor das Barrocas, 38, Aveiro, 3800–075, Portugal
[c] CITTA & University of Coimbra, R. Luís Reis Santos, Coimbra, 3030-788, Portugal

ARTICLE INFO

Keywords:
Emissions monitoring
Data privacy
Artificial intelligence
Sustainable cities and communities
Multimodal transportation

ABSTRACT

Cities worldwide have agreed on ambitious goals regarding carbon neutrality. To do so, policymakers seek ways to foster smarter and cleaner transportation solutions. However, citizens lack awareness of their carbon footprint and of greener mobility alternatives such as public transports. With this, three main challenges emerge: *(i)* increase users' awareness regarding their carbon footprint, *(ii)* provide personalized recommendations and incentives for using sustainable transportation alternatives and, *(iii)* guarantee that any personal data collected from the user is kept private.

This paper addresses these challenges by proposing a new methodology. Created under the FranchetAI project, the methodology combines federated Artificial Intelligence (AI) and Greenhouse Gas (GHG) estimation models to calculate the carbon footprint of users when choosing different transportation modes (*e.g.*, foot, car, bus). Through a mobile application that keeps the privacy of users' personal information, the project aims at providing detailed reports to inform citizens about their impact on the environment, and an incentive program to promote the usage of more sustainable mobility alternatives.

1. Introduction

According to the World Economic Forum (WEF) [1], *"mobility is a fundamental human need and an essential enabler of prosperity, but the current mobility paradigm is not sustainable"*. Quoting WEF, car travel causes millions of deaths every year, with a significant amount of Greenhouse Gas (GHG) emissions, and traffic congestion causing heavy financial loss. The European Commission also acknowledges that transportation is the leading cause of air pollution in cities [2]. With this, cities worldwide have agreed on ambitious goals towards 2030 regarding GHG emissions and carbon neutrality. Given this agenda, the global mobility system is in its early stage of massive transformation. Namely, policymakers are seeking ways to foster smarter, cleaner, and more inclusive mobility. For this to be possible, we argue that one must consider three main challenges.

Carbon footprint awareness. Citizens are not aware of their carbon footprint when using different transportation modes (*e.g.,* walking, bicycle, motorcycle, car, bus). Mobility patterns should be collected and leveraged to provide citizens with information about their impact on the environment.

Sustainable mobility. Cities are unable to implement greener strategies for active and shared mobility due to public transportation's low attractiveness. Citizens should be provided with personalized recommendations and incentives for using this type of transport.

Data privacy. Citizens should be aware of how their data is collected and used and, more importantly, that their personal information (*e.g.*, location) is kept private. Without this, the adoption of applications and services that target the two previous challenges will be limited.

To address all the challenges, this paper proposes a new methodology, created under the FranchetAI project [3], that promotes personalized and sustainable mobility behavior while preserving data privacy and increasing user trustworthiness. FranchetAI's methodology is built on top of the following pillars: *(i)* Artificial Intelligence (AI) and GHG estimation models to detect the type of transportation being used by a given citizen, along with its corresponding carbon footprint; *(ii)* state-of-the-art mechanisms that safeguard data collected from users' mobile devices by not sharing private/sensitive data with any external service (*e.g.*, cloud provider); *(iii)* compliance with European best practices in usability, accessibility, and explainable AI to clarify in an understandable way how users' data is being processed; and, finally, *(iv)* building

* Corresponding author.
E-mail address: claudia.v.brito@inesctec.pt (C. Brito).

https://doi.org/10.1016/j.treng.2024.100237
Received 29 April 2023; Accepted 3 March 2024
Available online 27 December 2024

up on the experience of gamification and habit changing to promote incentives (*e.g.,* rewards, vouchers, among others) to encourage the community to opt for sustainable mobility choices, as well as to create more awareness about sustainability among citizens. As output, users will be informed about their mobility choices' carbon footprint through *carbon digest reports* (daily and weekly) via a mobile application.

In more detail, AI models are used to determine the transportation mode being used by each user. Models are firstly trained with the open-source GeoLife dataset [4]. Then, these models are iteratively retrained with Federated Learning (FL) on users' mobile devices, using sensor data (*e.g.,* GPS/GNSS, accelerometer) collected through a mobile application. Although collected data never leaves users' premises, at each training iteration, the parameters (*i.e.,* model's gradients) are broadcasted to a centralized external server. Since the collected data and these broadcasted parameters contain sensitive user information, we rely on Differential Privacy (DP) to ensure the privacy of our solution. DP introduces random noise to the users' data and/or the broadcasted parameters allowing the model to not remember the data that it was trained with, thus not leaking any personal information when being queried by third-party entities (*e.g.,* municipalities and/or cities).

Further, by using explainable AI during the training stages, the methodology follows a user-centric approach to make data understandable to various stakeholders (*i.e.*, commuters, municipalities, and transportation operators). Namely, it allows understanding the impact of each data feature on the trained model and the inference result.

With the trained models one can then infer the type of transport for each user and use the result as input for a GHG model that estimates the carbon footprint of a given trip. This second model provides the necessary information to create the carbon digest above-mentioned.

Initial results comparing Decision Trees, Random Forest, Logistic Regression, and XGBoost algorithms, show the impact of different features on the prediction of the user's mode of transport. Namely, we show that it is possible to obtain results with over 80% accuracy when considering the distance and mean velocity of users' trajectories. Such results form the basis for training with more complex algorithms based on neural networks.

The end goal of the proposed methodology and proof-of-concept mobile application is to engage commuters to take greener options via gamification mechanisms. This is achieved by first comparing individual reports with the averages of local and European communities, as well as with the necessary targets to minimize climate change. Such comparison is the basis to provide rewards/incentives via local challenges promoted and funded by Non-Governmental Organizations (NGOs), decision-makers, and businesses willing to invest in carbon reduction. By nurturing sustainable travel behaviors and changing commuters' habits, this methodology aims to contribute to the direct reduction of CO2 emissions.

2. Literature review

Knowing how users interact on a daily basis with transportation utilities allows understanding their impact on the environment. Solutions targeting such a goal must be able to collect and determine what modes of transportation are being used while pinpointing their carbon footprint. This often requires collecting sensitive personal information from devices, such as mobile phones [5]. Therefore, one must also ensure that users' data is not disclosed to unwanted third-parties. With this, state-of-the-art solutions can be decomposed into three main subgroups: *(i)* mobility patterns and transportation mode, *(ii)* carbon footprint assessment, and *(iii)* federated learning for mobility.

2.1. Mobility patterns and transportation modes

Mobility pattern studies typically focus on how urban transportation is used by citizens and how it might be enhanced to better serve their needs and those of communities. Data collected from several sources (*e.g.*, GPS traces, weather conditions, traffic status) can be leveraged to deliver solutions with different purposes [6].

Currently, mobility patterns are classified into two main types: prediction and generation. Regarding mobility pattern prediction, studies have focused on:

Next-location prediction. It aims at understanding the future locations of users based on their previous behavior and historical mobility data. Such information can be beneficial for improving public health, reducing traffic congestion, and providing better travel recommendations. Additionally, it can help urban and public transportation planning, as cities can leverage this information to upgrade their road systems, urban infrastructure, and public transportation systems [5,7].

Crowd flow prediction. It assists cities in pinpointing areas of possible traffic congestion and infrastructural improvement based on crowd behavior. Once more, these studies are based directly on users' information and how they move around cities. However, they leverage the flow of the crowd itself and do not intend to predict an individual's next geographical position [5,7].

Mobility pattern generation studies also follow two different approaches:

Trajectory generation. This approach is based on individual GPS traces and trajectories, and the goal is to generate synthetic trajectories that follow distinct traveled distances and predictable human mobility patterns. The generation of these trajectories may aid in urban planning and avoids collecting citizens' geographical positions [5,8].

Flow generation. Based on the exact location of users, this approach gathers crowd flow information for a specific geographical region. This is a crucial process for transport planning and epidemic spread patterns [5,8].

The prediction and classification of mobility patterns can also be leveraged for defining the transportation mode of users. For instance, recent studies have used GPS traces for transportation-mode detection, classification, and prediction. These studies have shown high confidence in defining several classes of transportation (*e.g.*, walk, run, bike, car, bus, train) by using Machine and Deep Learning algorithms (ML/DL) [9–11]. Other solutions explore the use of multiple sensors to detect transportation modes alongside the GPS traces (*i.e.,* accelerometers, gyroscopes, and/or magnetometers).

Further, to improve the accuracy of these models, one can rely on additional information from public transportation networks and urban infrastructure (*e.g.*, roads, streets, highways, etc.). The interplay of these new variables allows enhancing the output of these prediction and classification models [9].

2.2. Carbon footprint assessment

Calculating and inferring the carbon footprint of an individual user is complex as it requires the collection of several data points and integration of distinct models [12].

Life Cycle Assessment (LCA) models allow the calculation of the carbon footprints for the entire life cycle of a specific transportation mode, from the manufacturing of the vehicle to its usage and maintenance and, finally, to its disposal (*cradle-to-grave* model). Such models provide a holistic view of vehicles' full life cycle and their impact on the environment [13,14].

On the other hand, the *Well-to-Wheel* (WtW) concept provides an alternative LCA model that focuses only on transport fuels and vehicles' usage. *Well-to-wheel* differs from *cradle-to-grave*, as it does not consider energy and emissions involved in building facilities and the vehicles, nor end-of-life aspects of the latter. The model analysis is often broken down into two stages entitled *well-to-tank* and *tank-to-wheel*. The first stage, known as *upstream stage*, incorporates fuel production, processing, and delivery, while the *downstream* stage deals with vehicle

operation. The WtW analysis is commonly used to assess total energy consumption or the energy conversion efficiency and emissions impact of different transport modes [15]. However, when used alone, these models are not sufficient to estimate a specific user's daily carbon footprint.

By knowing the transportation mode being used by a given user, and by combining information about the user's trajectory (*e.g.,* distance traveled, mean velocity, vehicle class, fuel type) with an LCA model, one can infer the carbon footprint of a commuter [16–18].

Alternatively, researchers are using ML and DL models to automate the estimation of GHG emissions. These models have achieved high accuracy rates and have been able to infer the carbon footprint of a user based solely on the transportation mode and traveled distance [19].

2.3. Federated learning for mobility

FL is a paradigm that allows ML models to be trained over distributed data without sharing it with third-party entities [20]. FL has been used in several domains such as healthcare, finance, and transportation. Specifically, in the transportation domain, FL has recently been used to train models that can predict the transportation mode of a user based on GPS traces [21,22].

While FL is based on the assumption that no user data is outsourced to third-party infrastructures, most algorithms do not contemplate malicious attacks that intend to infer sensitive data from the parameters or the trained model. To tackle such a challenge, privacy-preserving FL has emerged. Recent algorithms employ DP as a privacy measure to ensure that sensitive data and gradients are not disclosed to third-party entities [23]. Alongside, Multi-Party Computation (MPC) algorithms, namely secure aggregation, is an alternative algorithm applied to FL [24]. While DP trades off accuracy for privacy, MPC trades off performance for privacy.

Summary. In general, there is not yet a solution that comprises all the previous three main topics in a holistic way. The FranchetAI methodology aims to address this challenge. Specifically, in this paper, we focus on the detection of the transportation modes and the estimation of the carbon footprint of each individual citizen while promoting the preservation of users' data privacy. In contrast to the work proposed in this paper, current FL solutions leverage federated algorithms without the additional privacy measures above-mentioned. Moreover, these approaches do not extend further from the transportation mode classification and do not explore the calculation of GHG emissions.

3. Methodology

FranchetAI provides a digital rewarding solution for people opting for sustainable mobility options (*e.g.*, public transportation, electric vehicles), ensuring transparency and trustworthiness between the user and the different stakeholders creating the incentives. Such a solution aims to educate citizens about their carbon footprint while offering traveling alternatives and rewards for traveling more sustainably. The methodology proposed by FranchetAI and depicted in Fig. 1 has the end goal of reducing CO_2 emissions. To achieve this, one can split the methodology into several steps. First, a centralized server allows deploying a web platform that processes and visualizes, for instance, traffic flow, road topologies, and public transit networks (Fig. 1-❶). Second, a mobile application is deployed on each user's mobile device (Fig. 1-❷). This application collects itineraries from GPS, accelerometer, and gyroscope (Fig. 1-❸). It also allows users to input additional info on vehicle usage and parameters (*e.g.,* year, class, fuel type) (Fig. 1-❹).

Information collected by the mobile app and the web platform works as an input to a transportation model. This model, based on AI, infers users' transportation mode by using personal information from their trips and public information from public transit networks (Fig. 1-❺). The collected user data is also used for retraining the model to improve its accuracy. This is done in a privacy-preserving manner by resorting to federated learning and guaranteeing that private users' data remain on their premises (*i.e.,* their mobile devices).

The output of the previous model's inference is leveraged by an LCA model that estimates GHG emissions (Fig. 1-❻) and generates daily and weekly *carbon digest reports* (Fig. 1-❼). The confidence of this estimation depends on the users' pre-validation of the output of the transportation mode model. Finally, a sustainability model that receives as input the GHG estimation, intends to recommend alternative and more sustainable transportation solutions (Fig. 1-❽). This final model is based on the trajectories of each user's trip and on the currently available transportation solutions that lead to fewer or no emissions. To increase the engagement of users with such an application and with greener alternatives, the FranchetAI mobile app promotes incentives from local community businesses, NGOs, and cities (Fig. 1-❾). Finally, FranchetAI mobile application already has a proof-of-concept layout where it is possible to have a first glimpse of the proposed carbon digest.

Next, we further detail the transportation and emissions models, along with the concepts and tools underpinning these. Also, in Section 4.3 we show that these can achieve high accuracy, even when only considering users' data as input (*i.e*, without additional sources such as traffic congestion, and transport networks).

3.1. AI models

Our AI approach is decoupled into two main stages. First, we define the training dataset and how data is preprocessed. Then, we choose the ML models and DL architectures to train with the previous dataset, while defining also the FL framework and explainable tools to use.

Before diving into these stages, we briefly explain the common workflow of an FL system and how it can be used to train models on top of sensitive data. Further, we overview the concepts of differential privacy and explainable AI.

3.1.1. Federated learning

Following the enormous amounts of collected data from different sources, ML has become the *de facto* solution for analyzing and extracting insights from it. Nonetheless, new regulations (*e.g.*, GDPR) impose new approaches for analyzing data that may contain sensitive information.

FL has emerged as a new ML paradigm targeting Non-Independent and Identically Distributed (Non-IID) data [25,26], typically generated on the edge, local servers, and mobile devices. Specifically, FL is a type of distributed ML in which models are trained with data from different users, but sensitive information never leaves each user's premises.

In this setting (see Fig. 2), a centralized server has an initial trained model M_i and broadcasts M_i to every user (Fig. 2-❶). Typically, M_i is trained on previously collected, open-source, or private data, resulting in a collection of parameters and hyperparameters. On the user side, the device trains the model based on the user's data (Fig. 2-❷).

At each iteration, the centralized server asks N users for their new model parameters (Fig. 2-❸). Each user can define whether to participate or not in a given round and similarly, the centralized server can decline the parameters broadcasted from the decentralized users (Fig. 2-❹). At the end of this cycle, the server calculates the average of all obtained parameters (Fig. 2-❺) and broadcasts new parameters to every user, which updates locally its own model (Fig. 2-❻) [25,26].

3.1.2. Differential privacy

To further strengthen data privacy guarantees, FL may resort to MPC or DP as discussed in Section 2.3. In this work, we focus on the latter.

The intuition behind DP relies on the fact that changing any individual point on the input data will not change the query result but will limit the attacker's capabilities for deducing private information from data with high confidence [23]. To this end, DP anonymizes the

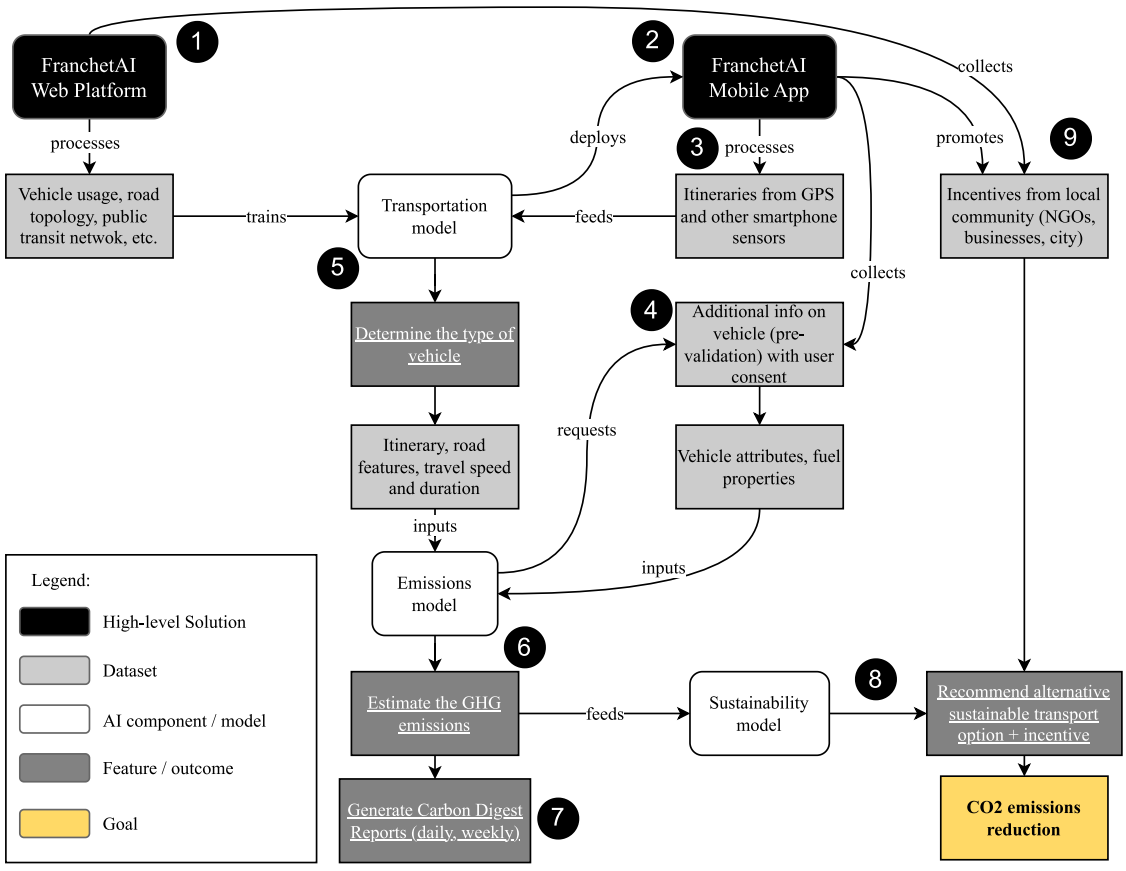


Fig. 1. Pipeline of the FranchetAI's methodology.

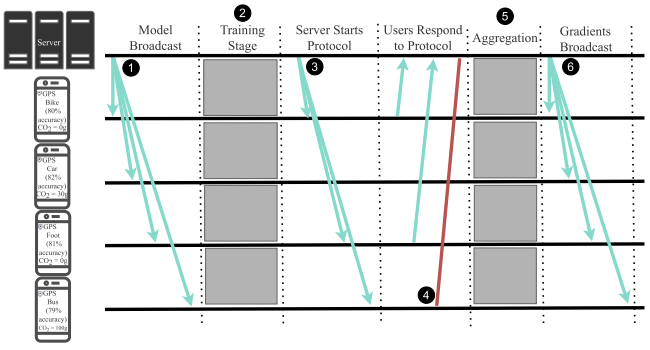


Fig. 2. FL protocol stages.

private information by introducing randomness or noise in the original data [23].

While the addition of noise to sensitive training data increases privacy-preserving guarantees, it also increases the models' error rate. In this sense, the trade-off between privacy and accuracy is a crucial aspect to consider when using DP [27]. To reduce the error rate, one can apply optimizations such as adaptive clipping [28], which continuously adapts the amount of noise introduced in each individual training sample.

In an FL setting, DP can be applied directly to users' data, or the parameters being broadcasted across users and the centralized server. In both cases, the goal is to guarantee that no private information is leaked when the trained model is queried by third-party entities (*e.g.,* cities, municipalities).

3.1.3. Explainable AI

Explainable AI (XAI) is a growing field in the AI community that promotes the transparency, interpretability, and trustworthiness of complex AI models. It aims to make models more understandable and interpretable for humans, and is built on top of three main goals [29]:

- **Model Interpretation:** The internals of the AI models are analyzed to understand how these calculate their decision outputs.
- **Model Visualization:** AI models are transformed into visual representations to understand their architectures and visualize the importance of each feature on the inference's result.
- **Model Explanation:** The inputs of the model are scored and sorted to understand their relevance to the trained model.

Currently, libraries, such as SHAP, Intel's Explainable AI Tools, or Google Cloud's XAI [30–32], allow the seamless integration of explainability tools during the training and inference stages of AI models.

In summary, XAI is a crucial component of the proposed methodology because it allows users to understand which data is leveraged for training the transportation model and how the final predictions are calculated.

3.1.4. Data preprocessing

To train the transportation model, we chose the GeoLife GPS Trajectories dataset [4]. It consists of 17,621 GPS trajectory data points from 178 users, each including latitude, longitude, and altitude information. Additionally, data points are labeled with different modes of transportation (classes). In this work, we focus on five classes, including vehicles (comprising individuals' cars and taxis), motorcycles, bikes, buses, and feet (comprising walking and running), which can be seamlessly changed if one wants to increase the number of classes.

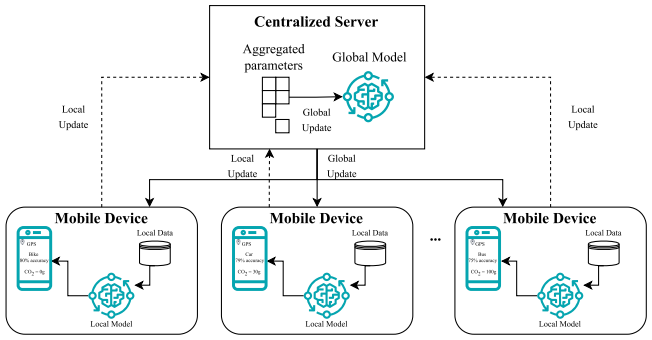


Fig. 3. Layout of the proposed FL system.

The preprocessing of input raw data is based on state-of-the-art methods and is used to calculate the velocity, acceleration, and distance of the trajectory for each user. The distance between two points is calculated using Geopy, which uses the geodesic distance to this end [33]. The velocity is calculated by dividing the distance between two points by the time taken to travel between these. Finally, the acceleration is calculated by dividing the velocity across two points with the time taken to travel between these.

Input data is then split into training and testing sets. The training set is used to train the model, whereas the testing set is used to evaluate the model. The split is performed using the `train_test_split` function from the SciKit-Learn library [34], using a 70/30 ratio, where 70% of the data is used for training and 30% for testing.

The preprocessed dataset is used to train a first version of the model in a centralized setting, which is then retrained with the data collected on the mobile device of each user by resorting to FL. The information collected at each mobile device is identical to the one reported in the GeoLife dataset and goes through the same preprocessing method.

3.1.5. Model development

Different ML and DL models were chosen to better understand their trade-offs between accuracy and execution time. For the ML models, we resorted to Random Forest, Decision Trees, Logistic Regression, and XGBoost, which were provided by the SciKit-Learn library [34]. The DL models were built based on the assumption that training data is provided through CSVs stored in a time-series database. As so, the models were developed with dense and long-short term layers. The literature supports the use of these architectures for similar time-series use cases [21]. Finally, these models were implemented on top of Tensorflow [35].

As explained previously and depicted in Fig. 3, all these models were firstly trained in a centralized fashion (*i.e.*, in a centralized server with the Geolife's dataset) and then, retrained in a federated setting (*i.e.*, across users' mobile devices with local data). For the latter setting, we started by using the PySyft framework [36,37]. However, this framework lacks support for mobile settings, which led to exploring two other alternatives, namely Tensorflow Lite [35] and Flower [38]. We opted for the latter since it supports straightforward integration with both the ML and DL models, and corresponding frameworks, used in the paper. Also, Flower allows federated training both in mobile settings and across data silos. This feature is important if cities provide access to other information, such as road topology and traffic congestion, that can be used along with users' data to improve the final models.

Flower also provides several averaging algorithms which comprise DP alternatives. These can be used to improve our solution's privacy-preserving guarantees and trustworthiness. Namely, we used the `DP-FedAvg` strategy to introduce random noise on the users' data, mitigating the leakage of their personal information (*e.g.*, location).

Finally, we used the SHAP library [30] to explain the models through Shapley values [39], which provide a way to measure the contribution of each feature to the model's final prediction.

3.2. GHG model

The methodology adopted for estimating Greenhouse Gases (GHG) and air pollution emissions is based on the *tank-to-wheel* stage of WtW LCA model, which only considers the operation of the vehicle [40].

In more detail, emissions are estimated based on the CORe INventory AIR emissions (CORINAIR) system, *i.e.*, the method approved by the European Environment Agency (EEA). CORINAIR adheres to the Intergovernmental Panel for Climate Change (IPCC) guidelines [41] used globally by environmental protection agencies for national and regional evaluations.

According to the IPCC Guidelines for greenhouse gasses, a compiler from the CORINAIR system builds a decision tree to select the appropriate methodology with different complexities and data requirements. As input to the compiler, we apply the Tier 3 methodology from EMEP/EEA emission inventory guidebook [42]. Moreover, the GHG estimations are based on the ultimate CO_2 emissions, which result from different processes (*i.e.,* combustion of fuel, combustion of lubricant oil, and addition of carbon-containing additives in the exhaust). This results in Eq. (1):

$$E_{ik} = e_{ik}(v) \cdot a_k, \tag{1}$$

where E_{ik} is the exhaust emissions of pollutant i induced by a vehicle technology k (in grams); e_{ik} is the emission factor as a function of the vehicle driving speed (in grams per kilometer); a_k is the transport activity in vehicle kilometers traveled for vehicle technology k.

The emissions are calculated individually for each user of the FranchetAI mobile application by considering the average driving speed of the road links that constitute an individual trip. The previous AI model(s) provides the information on traffic data (transportation mode, trip route, distance, and traveling speed) necessary to calculate clients' emissions from traffic activity. Also, information on vehicle technology is required, *i.e.*, the Euro Standard information, accessed by considering the age of the vehicle. Therefore, a user is asked to give this detailed information; otherwise, a default technology is used *(Euro 4)*.

4. Results

In this section, we highlight the main results of the paper. First, we assess the accuracy and training times of different AI models when trained with the GeoLife dataset in a centralized setting. Then, we showcase the viability of using our DL model within an FL setup. Then we present an output example for our emissions model, and finally, we overview an initial mock-up and layout of the FranchetAI mobile application.

4.1. AI models

The first tests with the GeoLife dataset helped limit and create general classes for the transportation modes. By limiting the number of classes, we were able to reduce the tree length and improve the results of Random Forests (RF) and Decision Trees (DT) models by around 5%, reaching an accuracy of 81% (as depicted in Fig. 4) in less than 3 min of training.

Plus, the implementation of Logistic Regression (LG) with and without cross-validation (LGCV) and XGBoost have shown lower accuracy results, namely 45% for both the logistic algorithms and around 70% for XGBoost. Regarding training times, the convergence of the algorithms took 1 min and 14.7 h, for LG and LGCV respectively, and 1.4 h for XGBoost.

Moreover, the developed DL model based on dense and long-short-term memory layers (NN) showed lower accuracy when compared to RF and DT. Yet, its implementation was able to obtain an accuracy of 75% with an execution time of 5 min.

These tests also allowed perceiving the impact of different input features. Namely, the first batch of tests was performed with the

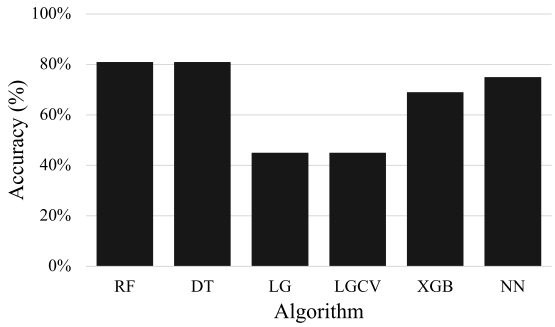


Fig. 4. Accuracy results for the tested models.

Fig. 5. Example of the current explainability outcome.

traveled distance, the mean velocity, and the mean acceleration in mind. However, the mean acceleration did not impact positively or negatively the accuracy of the trained models.

To assess the viability of using an FL setup, we sharded the test dataset into ten shards, each attributed to a distinct user's device (mobile application). These experiments focused only on the NN model. Initial results are promising, showing that the model can label the remaining trajectories (*i.e.*, test dataset) with similar accuracy. Also, the retraining of the model on the test dataset was performed in 5 rounds, in which the users' devices trained for 5 iterations and broadcasted their parameters to the centralized server. On the user's side, the differential privacy strategy used a constant value of clipping of 0.1 and a noise multiplier of 0.4. In the end, a global model was saved containing the aggregation of all the parameters and kept the initial accuracy of 75%.

The explainability tool allowed us to understand the weight given to each feature per class (Fig. 5). With classes (*e.g.*, car, foot, bus) being the output (of the model), the mean velocity and the distance of each trajectory are the features of the model. For instance, for *class 4*, the velocity feature is more relevant than distance.

Although the evaluated solution presents a viable option for mobile settings, other features such as instant velocity and accelerometer and/or gyroscope information, which can be collected through the mobile application, should be made available to train new models and check whether these can improve training speed and the model's accuracy.

4.2. GHG estimation

Our current emissions model requires the following information: (i) mean velocity of the trajectory; (ii) type of fuel of the car; (iii) category of the vehicle (*i.e.*, passenger, bus, heavy duty and, motorcycle); (iv) the total distance of the trip; and (v) the year of the vehicle. The year and category are further used to define the Euro Standard. Still, when the user does not disclose such information, the GHG emissions are estimated based on Euro 4, while the previously trained AI models define the vehicle category.

Fig. 6 presents an example of the usage of our emissions model. For instance, for a user commuting for 30 min at 30 km/h, totaling a distance of 15 Km, one may analyze the CO_2 induced by different

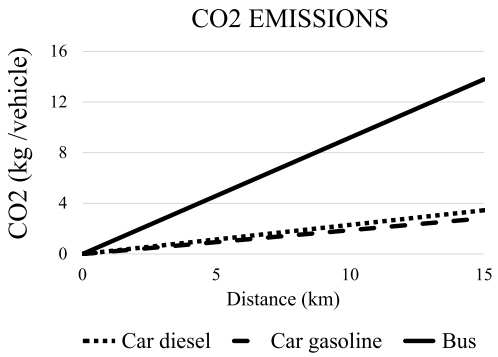


Fig. 6. CO_2 emissions of a user commuting with a diesel car, gasoline car and bus.

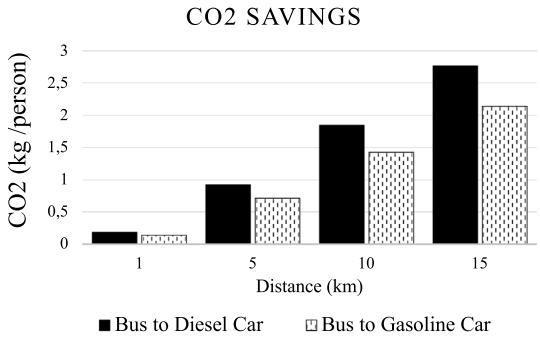


Fig. 7. CO_2 savings of a user commuting with a bus instead of using a car fueled by diesel or gasoline.

vehicle modes (*e.g.*, diesel/gasoline car or bus) in kilograms per vehicle. Also, by assuming that an urban bus will have an occupancy of 20 people, the impact of using such a more environmentally friendly vehicle is presented in Fig. 7.

4.3. FranchetAI mobile prototype

As previously mentioned, the FranchetAI mobile application (see Fig. 8) aims to offer a digital rewarding mechanism for users opting for more sustainable mobility solutions. With direct access to a daily and weekly carbon digest, the citizens become aware of their own carbon footprint and how they compare to others from the same community or globally. While the application collects the needed information for the previous models to work, this data is stored locally and leveraged only inside the users' mobile devices.

Furthermore, FranchetAI helps increase the users' awareness of how their transportation choices impact the environment while rewarding them for *good behavior* through incentives provided, for example, by local stores and services. Therefore, FranchetAI plays a crucial role in achieving the Sustainable Development Goals regarding climate change and helps build cities' economies while promoting local businesses. This prototype must be fully implemented and deployed into the pilot stage for a complete proof-of-concept solution. With this, the focus will be on the young adult generations, who are typically more prompt to test new environment-aware solutions. This stage will also allow us to try and improve the feasibility of such a novel solution.

5. Conclusion

The proposed methodology leverages best practices for engaging citizens in taking more conscious environmental decisions while ensuring their right to privacy. Specifically, the solution uses AI models to

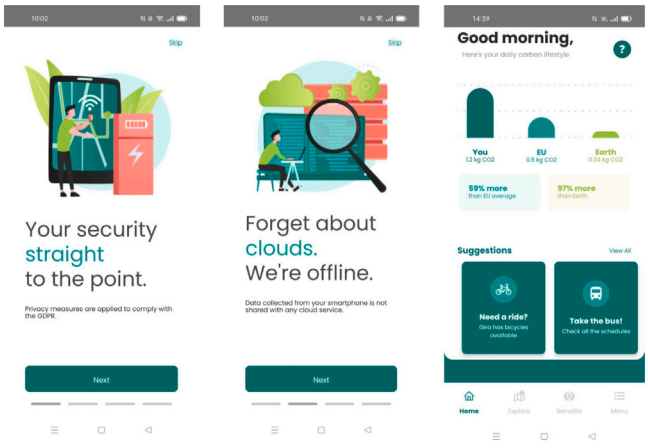


Fig. 8. FranchetAI mobile application.

detect the transportation modes of users while combining such information with a GHG emissions model to provide recommendations and incentives to change users' habits. Importantly, users have complete control of their data, knowing which data is used for locally inferring the system's models and which is used for training new models. Further, by following a federated learning setting and security protocols, users' sensitive information is never disclosed to unwanted third-parties.

Regarding future work, the methodology will be validated within real-world scenarios. Data from past and ongoing initiatives (namely other R&D projects) is being used as well as open data and third-party platforms to ensure the solution is as *off-the-shelf* as possible (although local context and data will help personalize it to the target communities). Also, we want to evaluate other strategies for the federated system and understand the trade-off between accuracy and privacy when adding differential privacy.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The FranchetAI project is part of the AI4 Cities project that has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 871914. This work was also supported by the Portuguese Foundation for Science and Technology through PhD Fellowships SFRH-BD-146528-2019 (Cláudia Brito) and DFA/BD/5881/2020 (Tânia Esteves).

Data availability

Data will be made available on request.

References

[1] WEF (World Economic Forum, Shaping the future of mobility, 2022, https://www.weforum.org/centres-and-platforms/shaping-the-future-of-mobility.
[2] EC (European Commission), Transport emissions, 2016, https://climate.ec.europa.eu/eu-action/transport-emissions_en.
[3] FranchetAI, Absorbing traffic pollution with AI, 2022, http://franchet.ai.
[4] Yu Zheng, Xing Xie, Wei-Ying Ma, et al., GeoLife: A collaborative social networking service among user, location and trajectory., IEEE Data Eng. Bull. 33 (2) (2010) 32–39.
[5] Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, Luca Pappalardo, A survey on deep learning for human mobility, ACM Comput. Surv. 55 (1) (2021) 1–44.
[6] Roelant A. Stegmann, Indrė Žliobaitė, Tuukka Tolvanen, Jaakko Hollmén, Jesse Read, A survey of evaluation methods for personal route and destination prediction from mobility traces, Wiley Interdisciplinary Rev.: Data Min. Knowl. Discov. 8 (2) (2018) e1237.
[7] Jonathan Ayebakuro Orama, Assumpció Huertas, Joan Borràs, Antonio Moreno, Salvador Anton Clavé, Identification of mobility patterns of clusters of city visitors: an application of artificial intelligence techniques to social media data, Appl. Sci. 12 (12) (2022) 5834.
[8] Yahua Zhang, Anming Zhang, Jiaoe Wang, Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China, Transp. Policy 94 (2020) 34–42.
[9] Haosheng Huang, Yi Cheng, Robert Weibel, Transport mode detection based on mobile phone network data: A systematic review, Transp. Res. C 101 (2019) 297–312.
[10] Lin Wang, Hristijan Gjoreski, Mathias Ciliberto, Sami Mekki, Stefan Valentin, Daniel Roggen, Enabling reproducible research in sensor-based transportation mode recognition with the Sussex-Huawei dataset, IEEE Access 7 (2019) 10870–10891.
[11] Xiaoyuan Liang, Yuchuan Zhang, Guiling Wang, Songhua Xu, A deep learning model for transportation mode detection based on smartphone sensing data, IEEE Trans. Intell. Transp. Syst. 21 (12) (2019) 5223–5235.
[12] Fahimeh Golbabaei, Tan Yigitcanlar, Jonathan Bunker, The role of shared autonomous vehicle systems in delivering smart urban mobility: A systematic review of the literature, Int. J. Sustain. Transp. 15 (10) (2021) 731–748.
[13] Christian Spreafico, Davide Russo, Exploiting the scientific literature for performing life cycle assessment about transportation, Sustainability 12 (18) (2020) 7548.
[14] Szczurowski Jakub, Lubecki Adrian, Balys Mieczyslaw, Brodawka Ewelina, Zarkbska Katarzyna, Life cycle assessment study on the public transport bus fleet electrification in the context of sustainable urban development strategy, Sci. Total Environ. 824 (2022) 153872.
[15] Shrey Verma, Gaurav Dwivedi, Puneet Verma, Life cycle assessment of electric vehicles in comparison to combustion engine vehicles: A review, Mater. Today: Proc. 49 (2022) 217–222.
[16] João C. Ferreira, Vítor Monteiro, José A. Afonso, João L. Afonso, Tracking users mobility patterns towards CO 2 footprint, in: Distributed Computing and Artificial Intelligence, 13th International Conference, Springer, 2016, pp. 87–96.
[17] Oana Lorintiu, Andrea Vassilev, Transportation mode recognition based on smartphone embedded sensors for carbon footprint estimation, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems, ITSC, IEEE, 2016, pp. 1976–1981.
[18] Maria Kugler, Sebastian Osswald, Christopher Frank, Markus Lienkamp, Mobility tracking system for CO2 footprint determination, in: Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, 2014, pp. 1–8.
[19] Wenxiang Li, Yuanyuan Li, Ziyuan Pu, Long Cheng, Lei Wang, Linchuan Yang, Revealing the real-world CO2 emission reduction of ridesplitting and its determinants based on machine learning, 2022, arXiv preprint arXiv:2204.00777.
[20] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, Blaise Aguera y Arcas, Communication-efficient learning of deep networks from decentralized data, in: Artificial Intelligence and Statistics, PMLR, 2017, pp. 1273–1282.
[21] Iago C. Cavalcante, Rodolfo I. Meneguette, Renato H. Torres, Leandro Y. Mano, Vinícius P. Gonçalves, Jó Ueyama, Gustavo Pessin, Georges D. Amvame Nze, Geraldo P. Rocha Filho, Federated system for transport mode detection, Energies 15 (23) (2022) 9256.
[22] Fuxun Yu, Zirui Xu, Zhuwei Qin, Xiang Chen, Privacy-preserving federated learning for transportation mode prediction based on personal mobility data, High-Confid. Comput. 2 (4) (2022) 100082.
[23] Cynthia Dwork, Aaron Roth, et al., The algorithmic foundations of differential privacy, Found. Trends Theor. Comput. Sci. 9 (3–4) (2014) 211–407.
[24] Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, Karn Seth, Practical secure aggregation for federated learning on user-held data, 2016, arXiv preprint arXiv:1611.04482.
[25] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konečnỳ, Stefano Mazzocchi, Brendan McMahan, et al., Towards federated learning at scale: System design, Proc. Mach. Learn. Syst. 1 (2019) 374–388.
[26] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, Virginia Smith, Federated learning: Challenges, methods, and future directions, IEEE Signal Process. Mag. 37 (3) (2020) 50–60.
[27] H. Brendan McMahan, Daniel Ramage, Kunal Talwar, Li Zhang, Learning differentially private recurrent language models, 2017, arXiv preprint arXiv:1710.06963.
[28] Galen Andrew, Om Thakkar, Brendan McMahan, Swaroop Ramaswamy, Differentially private learning with adaptive clipping, Adv. Neural Inf. Process. Syst. 34 (2021) 17455–17466.
[29] Wojciech Samek, Thomas Wiegand, Klaus-Robert Müller, Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models, 2017, arXiv preprint arXiv:1708.08296.

[30] Scott M. Lundberg, Su-In Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Process. Syst. 30 (2017).
[31] Intel, Intel® explainable AI tools, 2023, https://www.intel.com/content/www/us/en/developer/articles/reference-implementation/explainable-ai-tools.html. (Accessed 28 April 2023).
[32] Google Cloud, Explainable AI, 2023, https://cloud.google.com/explainable-ai. (Accessed 28 April 2023).
[33] GeoPy, Welcome to GeoPy's documentation! — GeoPy 2.3.0 documentation, 2022, https://geopy.readthedocs.io/en/stable/#module-geopy.distance. (Accessed 21 April 2023).
[34] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al., Scikit-learn: Machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[35] TensorFlow, TensorFlow lite | ML for mobile and edge devices, 2023, https://www.tensorflow.org/lite.
[36] Theo Ryffel, Andrew Trask, Morten Dahl, Bobby Wagner, Jason Mancuso, Daniel Rueckert, Jonathan Passerat-Palmbach, A generic framework for privacy preserving deep learning, 2018, arXiv preprint arXiv:1811.04017.
[37] Alexander Ziller, Andrew Trask, Antonio Lopardo, Benjamin Szymkow, Bobby Wagner, Emma Bluemke, Jean-Mickael Nounahon, Jonathan Passerat-Palmbach, Kritika Prakash, Nick Rose, et al., Pysyft: A library for easy federated learning, in: Federated Learning Systems: Towards Next-Generation AI, Springer, 2021, pp. 111–139.
[38] Daniel J. Beutel, Taner Topal, Akhil Mathur, Xinchi Qiu, Titouan Parcollet, Pedro P.B. de Gusmão, Nicholas D. Lane, Flower: A friendly federated learning research framework, 2020, arXiv preprint arXiv:2007.14390.
[39] Mukund Sundararajan, Amir Najmi, The many Shapley values for model explanation, in: International Conference on Machine Learning, PMLR, 2020, pp. 9269–9278.
[40] S. Gupta, V. Patil, M. Himabindu, R.V. Ravikrishna, Life-cycle analysis of energy and greenhouse gas emissions of automotive fuels in India: Part 1–Tank-to-Wheel analysis, Energy 96 (2016) 684–698.
[41] IPO Change, IPCC Guidelines for National Greenhouse Gas Inventories, Institute for Global Environmental Strategies, Hayama, Kanagawa, Japan, 2006.
[42] EMEP/EEA, Exhaust emissions from road transport. Passenger cars, light-duty trucks, heavy-duty vehicles including buses and motor cycles, 13/2019, 2019, European Monitoring and Evaluation Programme (EMEP), Air Pollutant Emission Inventory Guidebook 2019.