Emission-Aware Federated Learning: A Case Study on Transportation and Carbon Footprint

Cláudia Brito, Noela Pina⋆†, Ricardo Vitorino⋆, Inês Cunha‡, João Paulo
INESC TEC & University of Minho, ⋆Ubiwhere, †CITTA & University of Coimbra, ‡Unaffiliated

1 INTRODUCTION

Cities worldwide have agreed to ambitious goals regarding carbon neutrality and GHG emissions. Based on this ambitious roadmap, smart cities face challenges regarding active and shared mobility owing to public transportation's low attractiveness and lack of real-time multimodal information for citizens (that integrates public transport). These struggles have increased the community's lack of awareness of their mobility choices' carbon footprints and low motivation to change habits.

Although there have been efforts to offer solutions to raise awareness among citizens, individual perceptions can be different. Citizens leave their carbon footprint depending on their transportation and mobility patterns. Therefore, an increasing need exists to leverage an individual's data to promote a user-specific solution. Nonetheless, many consumers are reluctant to use third-party solutions that may compromise their personal or private data because of their lack of privacy guarantees.

With this, our solution is built on top of three main pillars, namely: *(i)* state-of-the-art mechanisms to safeguard data collected from smartphones by not sharing private/sensitive data with any cloud service [3, 5]; *(ii)* compliance with European best practices in usability, accessibility and explainable AI to clarify in an understandable way how the data is being processed and how the results are achieved, and, finally, *(iii)* building up the experience on gamification and habit changing to promote incentives (rewards, vouchers, among others) to encourage the community to opt for sustainable trip choices as well as to create more awareness.

This study proposes a solution that promotes personalized and sustainable travel behaviors while preserving data privacy and increasing user trustworthiness. Focusing on the practical usage of Federated Learning (FL) in fully distributed environments, Explainable AI (XAI), and mobile environments, this solution is built upon users' data to provide users with global knowledge about their carbon footprint.

Machine Learning (ML) models are built to predict user modal choice by considering standard sensor data from smartphone devices (GPS/GNSS, accelerometer, etc.). The predicted results are then used as input to a second model for GHG emissions, which estimates the users' carbon footprint.

2 DESIGN OVERVIEW

The design goal of this solution is to leverage ML models to understand users' mobility patterns and their chosen transportation modes. By following a FL distributed setting, it is intended that an ML model deployed locally in each user's mobile device infers and trains with the users' data. While training the models locally with each user's individual data, FL also assumes a global model in a centralized server that is updated regularly. Briefly, N users share their gradients with this centralized server which aggregates the shared information and sends a new batch of gradients for all connected users. The user's local model uses the received gradients in the next iteration. Moreover, privacy is guaranteed by resorting to Differential Privacy. Adding a predefined amount of noise to a user's data or parameters guarantees that such information cannot be disclosed. Nonetheless, finding the correct amount of noise for privacy-preserving data while maintaining the accuracy of each local model is challenging.

As inferred from the model, the transportation mode of each user is later used to calculate its carbon footprint and GHG emissions. This approach increases the user's awareness of their global emissions from transportation and promotes a change in the users' mobility patterns.

Even though privacy-preserving user data is one of the main focal points, users usually want to understand how their data is leveraged. Explainable AI stems from this requirement and allow users to understand which data is being used as well as what is the real impact on the trained models and the output of these models. The proof-of-concept solution was implemented on top of Flower [1] and SHAP [2] and the ML models were trained using GeoLife GPS trajectories [4].

3 FUTURE DIRECTIONS

Differential privacy has been proving itself as a resourceful solution for offering privacy-preserving solutions. Nonetheless, the trade-off between the models' accuracy and data privacy is a focal point for offering these solutions. A study understanding the correct point of balance and optimising the end results, is, therefore, an important future research direction. Also, the proposed proof-of-concept could profit from validation from real-world scenarios and the focus should rely on real-world datasets and specific communities that benefit from different transportation modes.

REFERENCES

[1] Daniel J Beutel, Taner Topal, Akhil Mathur, Xinchi Qiu, Titouan Parcollet, and Nicholas D Lane. 2020. Flower: A Friendly Federated Learning Research Framework. *arXiv preprint arXiv:2007.14390* (2020).

[2] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc.

[3] Dimitrios Michael Manias and Abdallah Shami. 2021. Making a case for federated learning in the internet of vehicles and intelligent transportation systems. *IEEE Network* (2021).

[4] Yu Zheng, Hao Fu, Xing Xie, Wei-Ying Ma, and Quannan Li. 2011. *Geolife GPS trajectory dataset - User Guide* (geolife gps trajectories 1.1 ed.). https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/

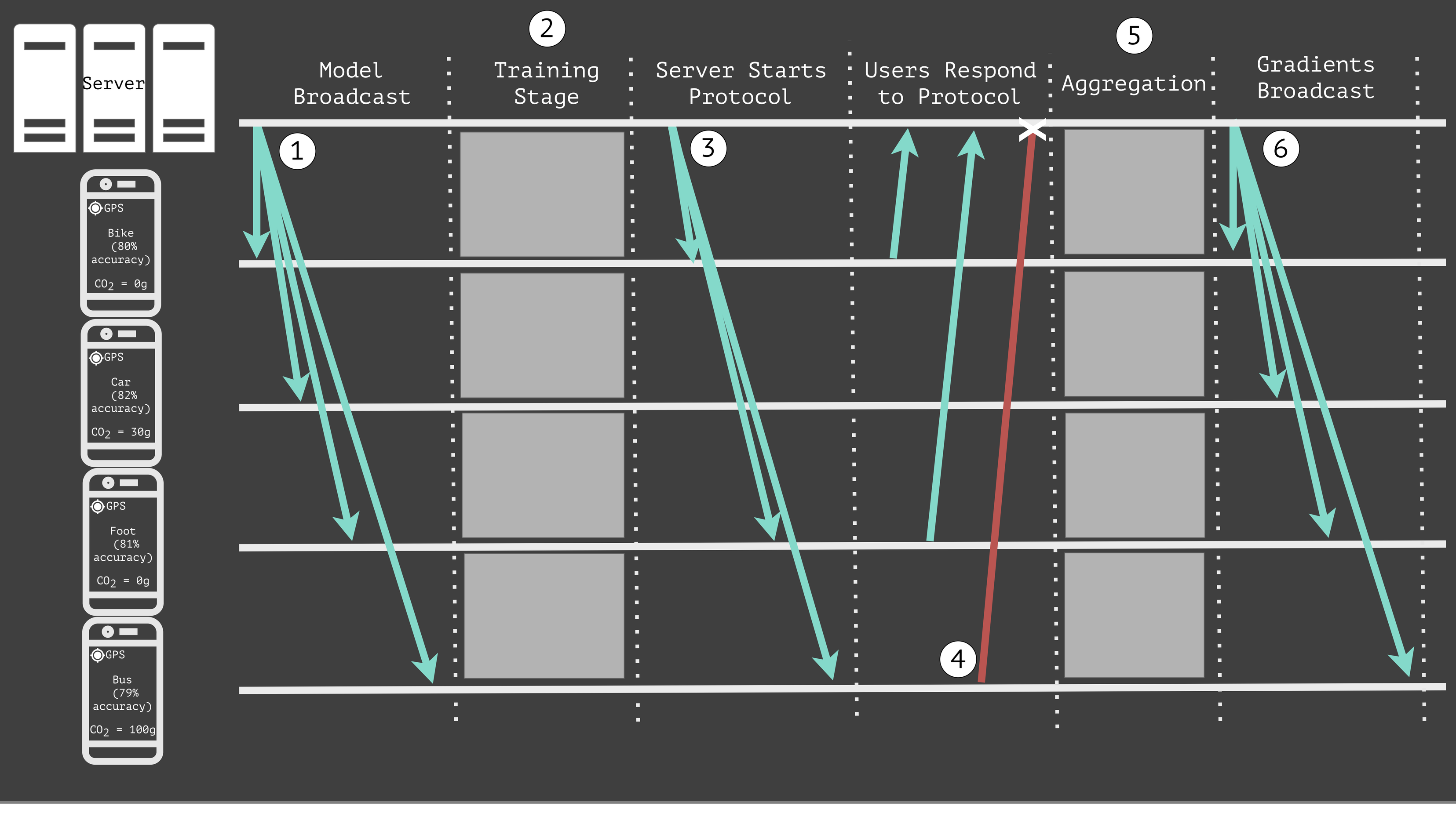
[5] Yuanshao Zhu, Shuyu Zhang, Yi Liu, Dusit Niyato, and JQ James. 2020. Robust federated learning approach for travel mode identification from non-IID GPS trajectories. In *IEEE 26th International Conference on Parallel and Distributed Systems (ICPADS)*.

Emission-Aware Federated Learning: A Case Study on Transportation and Carbon Footprint

Cláudia Brito★, *Noela Pina*♦*, *Ricardo Vitorino*♦, *Inês Cunha*+, *João Paulo*★

★ *INESC TEC & University of Minho*, ♦ *Ubiwhere*, * *CITTA & University of Coimbra*, + Unaffiliated

The addition of Federated Learning to Urban Transportation enables the promotion of **sustainable** and **personalised travel behaviours** while preserving **data privacy**.



1. Introduction

- Cities worldwide have ambitious goals regarding carbon neutrality.
- Users' lack awareness on their carbon footprint and the motivation to change habits.
- There is still reluctancy to use software tools due to possible leakage of users' private data.

2. Design

The main goal is to **preserve the privacy of users' data** while increasing **awareness** on their **carbon footprint**.

We propose a methodology that:

1. Detects and classifies transportation modes based on ML/DL models.
2. Estimates CO_2 emissions for each citizen (through a daily carbon digest).
3. Ensures the privacy of citizens data.
4. Integrates Explainable AI to make data and models understandable for citizens.

3. Results

- GeoLife dataset was used to train and create **general labels** for the transportation modes (e.g., car, foot, bus).
- The system was tested with **10 clients**, achieving around **75%** accuracy.

4. Future Directions

- Study the **trade-offs** between **data privacy** and **models accuracy** when applying differential privacy.
- Assess with **realistic scenarios** and **datasets**.

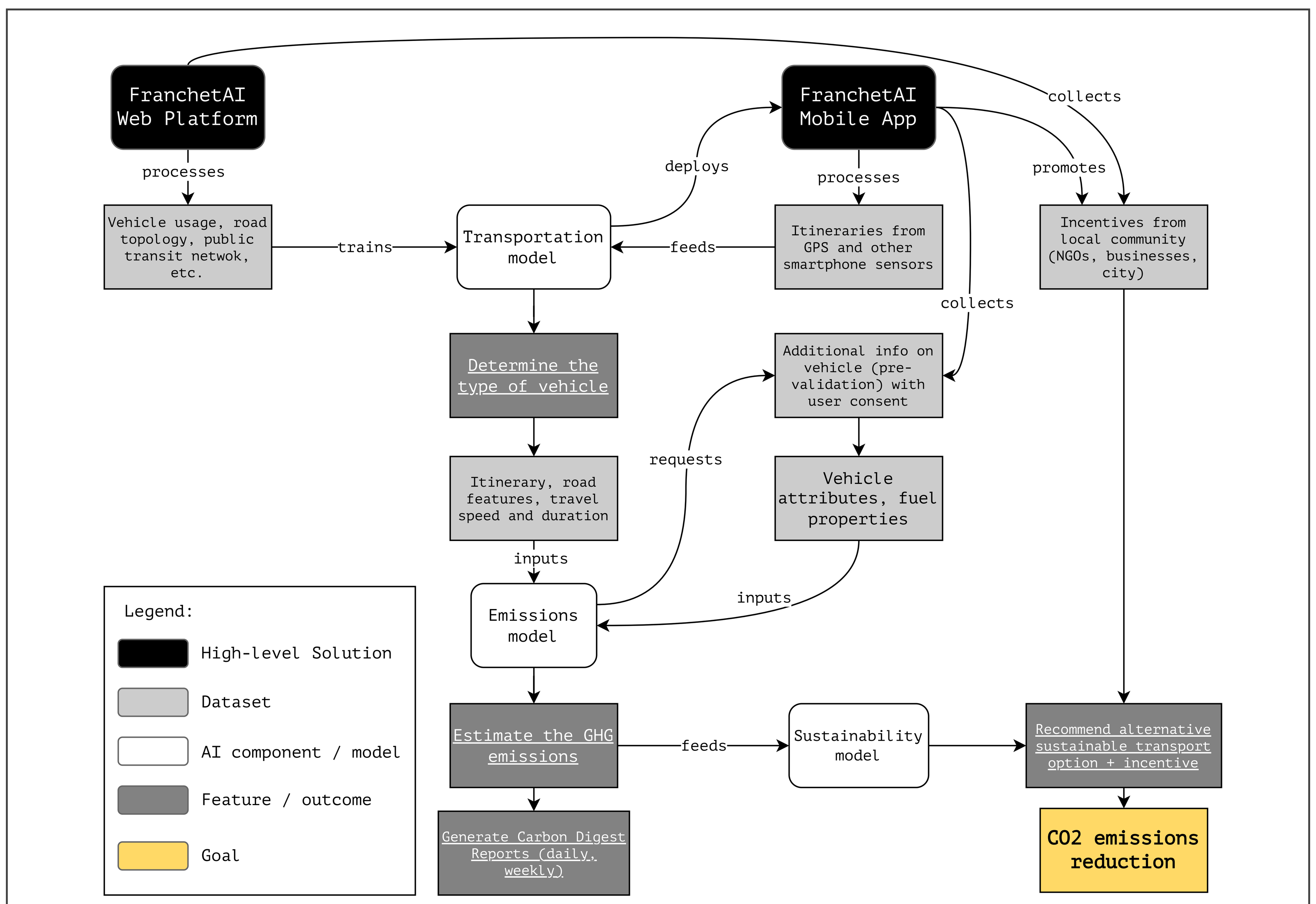


Figure 1. Pipeline of the full proposed methodology.

Figure 2. Explainability feature allows us to understand the weight given to each feature to label each class.

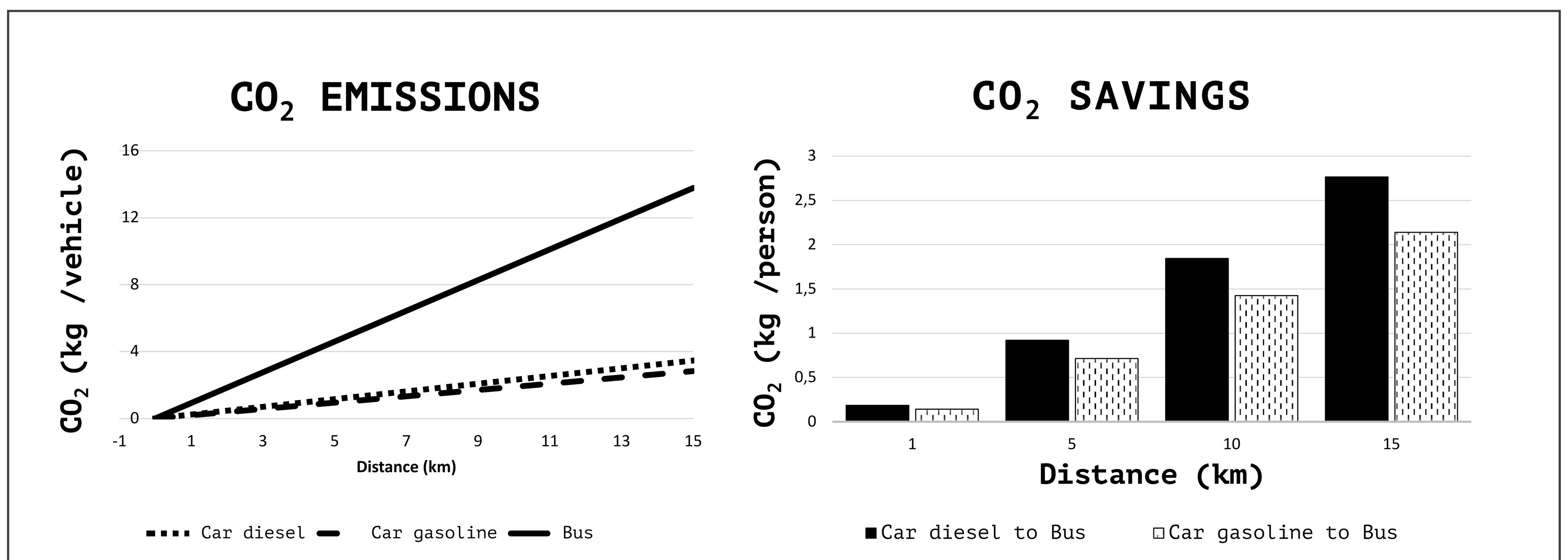


Figure 3. CO_2 emissions (left) and CO_2 savings (right) of a user commuting with a diesel car, gasoline car and a bus.

Acknowledgements

This work was carried out within the framework of AI4 Cities project (EU H2020, No 871914). This work was also supported by the Portuguese Foundation for Science and Technology through a Ph.D. Fellowship (SFRH/BD/146528/2019) and by the ERDF - European Regional Development Fund, through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme under the Portugal 2020 Partnership Agreement, and by National Funds through the FCT - Portuguese Foundation for Science and Technology, I.P. on the scope of the CMU Portugal Program within project AIDA, with reference POCI-01-0247-FEDER-045907.

claudia.v.brito@inesctec.pt