# Interactions Between Drivers and Road Infrastructure Characteristics: Combining OBD and Geographic Data to Classify Driver Behaviors with Feed-Forward Neural Network

V. Nicolosi[1], M. Mameli[2], S. Shiralizadeh[2], I. G. Coltea[2], Mauro D'Apuzzo[3], and G. Cappelli[1,3(✉)]

[1] Department of Enterprise Engineering "Mario Lucertini", University of Rome Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy
{nicolosi,cppgpp01}@uniroma2.it, giuseppe.cappelli1@unicas.it
[2] Key 2 Business S.r.l., via del Viale dell'Esperanto 71, 00133 Rome, Italy
{m.mameli,s.shiralizadeh,i.coltea}@key2.it
[3] University of Cassino and Southern Lazio, Via G. Di Biasio 43, 03043 Cassino, Italy
dapuzzo@unicas.it

**Abstract.** Driver behavior is the set of actions that a road user undertakes during a driving task. There is a huge interest in studying driver behaviors evaluating fuel consumption and improving safety. Using in-vehicle sensors is a widely adopted methodology to fulfill these objectives. The large amount of data generated from multiple sensors opens the doors to machine learning and deep learning algorithms which are more adequate than other methodologies. In this paper, a Feed Forward Neural Network is trained and tested with OBD and geographic data to classify driver behaviors. Several datasets are analyzed but the most adequate has resulted in the DDD20 dataset. This contains a larger amount of data than the other one, with 51 h and 4000 km of total driving times and distances. After the selection of the dataset and the enrichment with geographical data, feature selection, and data labeling techniques and algorithms are implemented. The model shows a high level of accuracy (above 98%) for the three classes of driver behavior studied.

**Keywords:** OBD · Geographic Data · Driver Behaviors · Road Safety

## 1 Introduction

The spread of new and effective computing technology, machine learning methods, and the introduction of Information and Communication Technology (ICT) in more digitalized vehicles has helped researchers discover and highlight the mechanisms behind driving behaviors [1]. The recent boost in the last decade could also be addressed by the growth and improvement of equipment and sensors used to collect data [2–4].

The literature [1, 5] reveals that the main sensors employed could be classified into three sets: the first is in-vehicle data recording systems that use CAN-bus and OBD

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026
O. Gervasi et al. (Eds.): ICCSA 2025 Workshops, LNCS 15892, pp. 51–64, 2026.
https://doi.org/10.1007/978-3-031-97638-4_4

interfaces; the second is smartphones and the different sensors on these devices; and finally, behavior detection methods that implement real-time methods, such as statistical and machine learning algorithms [6].

Once all the data are collected through sensors, they need to be processed. As the study of Elassad et al. [7] has shown, the most implemented machine learning methods are Neural Networks, Support Vector Machines, Ensemble Learners, Bayesian Learners, and Decision Trees.

The efforts in this field also aim to increase safety and improve fuel consumption. One of the most important principles of the Safe System Approach is that "Humans make Mistakes" when using the road system [8, 9]. Some authors suggest that 90% of road accidents are due to driver behaviors [10, 11]: so, understanding human behaviors on the road is a leading priority for improving overall safety. It is worth mentioning that often fuel consumption aspects are correlated with road safety. Prior research highlights how eco-driving styles lead to a double outcome: the first reduces fuel consumption and dampens unsafe driving behaviors [12–14].

But how safety and fuel consumption are related? Yen et al. [13] provide a clear framework for the problem. The main contributory factors affecting safety and fuel consumption, according to [13] are low and safe speeds, calm behaviors during driving, adequate route planning, and use of air conditioning in the cars. In the study of Graba et al. [15], the authors find that an increment of the average acceleration of 0.1 $m/s^2$ leads to an increment of 0.15 J $(kg\ m)^{-1}$ in unit total energy consumption. The study of [16] shows also that smooth acceleration without sudden speed fluctuations (typical of economical driving styles), could decrease fuel consumption up to 4.83 l/100 km. In contrast, a more dynamic style with frequent and higher accelerations could increase fuel consumption up to 6.15 l/100 km.

In this paper, a Feed Forward Neural Network is trained and tested with OBD and geographic data to classify driver behaviors. After the selection of the dataset and the enrichment with geographical data, feature selection, and data labeling techniques and algorithms are implemented. The model shows a high level of accuracy for the three classes of driver behavior studied.

To try to provide a clear framework in which this paper is located, after a brief introduction, in Sect. 2, a literature review of studies that aim to predict and classify driver behaviours is provided. In Sect. 3, a deep dive on the adopted methodology is provided with a particular focus on the dataset used to perform analysis, which features have been selected, how the data are labelled, and how OBD data are enriched with geographical information to perform feature-context analysis. In Sect. 4, the results obtained with the Feed Forward Neural Network are shown, and in the final section (Sect. 5), the conclusions, limitations, and future developments are drawn.

## 2 Related Works

According to Shahverdy et al. [17], driving behavior is the set of actions of users during driving tasks and could be classified as safe, aggressive, distracted, drowsy, and drunk driving. As suggested by the same authors [17], to categorize driver behavior, non-visual features of vehicles or drivers [18–20], such as vehicle speeds and accelerations, and

visual features, such as driver's eye movements [21, 22] could be useful to the task. It also emerges that the most common methods to classify driver behaviors are machine learning algorithms [17] that are implemented to handle such an amount of data from sensors and vehicle units.

Gear selection, braking, fuel injection, and other functions are monitored and optimized with sensors and electronic control units (ECUs) in vehicles [23]. All these control units can interact with each other through a bus, also called a controller area network (CAN), that allows one to read the signals from the sensors located in the vehicle and have an overview of driving performances [23, 24]. The on-board-diagnostic module (OBD) can get data from the ECUs, and so provide different parameters such as engine speeds, loads, and more [25, 26].

In the work of Ameen et al. [27], the authors use OBD interface to classify safe or not drive behavior from accelerations. The authors find that safe driving behavior is marked by values of acceleration of 2 $m/s^2$, while an increased risk of collision starts with more aggressive driving with acceleration values from 4 $m/s^2$. In addition to this classification, the results show an accuracy of 90% in predicting the driving scenario.

On the contrary, Abukhalil et al. [28] propose a method to evaluate fuel consumption by using the information system OBD-II, specifically the engine revolutions per minute (RPM) and throttle position sensor (TPS), with the help of Machine Learning (ML) Models. Several vehicles are used in field tests to evaluate fuel consumption. The model that is implemented to retrieve this relationship, relies on a Support Vector Machine classifier and a Lagrange interpolation, with reliable results.

The combined use of ML and OBD interfaces is also shown in the study of Campoverde et al. [29]. The authors heavily use these techniques: they implement random forest algorithms to make a variable importance analysis and then train an artificial neural network to estimate vehicle emissions. A K-means classifier is also obtained for the gear selection by the driver. In addition to that, four artificial neural networks are trained to evaluate pollutant emissions.

In their work, Lattanzi et al. [23] investigate the possibility of studying driving behavior using different data sources such as external and in-vehicle sensors (sensor fusion). A support vector classifier and artificial neural network are trained and tested to perform the classification task. The results show that the artificial neural network has an accuracy of 90% in classifying unsafe driver behavior correctly, while the support vector machine reaches 88%.

In another related paper by the same authors [30], only in-vehicle data are used, and more than 26 h of driving are recorded. In this study, a support vector machine and a feed-forward neural network are trained and tested. Both algorithms show a prediction accuracy higher than 90%.

In the work of Yen et al. [13], the authors used different deep learning classifiers, such as the Elman neural network, feed-forward neural network, and layer recursive neural network, to analyze fuel consumption. The proposed models predict real fuel consumption with an accuracy of over 96%.

In a recent work by Kumar and Jain [31], the authors trained and tested a support vector machine, Adaboost, and random forest to classify driver behaviors. The models

show a prediction accuracy of 99%, 99%, and 100%, respectively, higher than the other studies analyzed, with data collected only within the OBD.

This brief overview of related works shows a huge interest in studying driver behaviors and evaluating fuel consumption. Using in-vehicle sensors is a widely adopted methodology to fulfill these objectives. The large amount of data generated from multiple sensors opens the doors to machine learning and deep learning algorithms [1, 32], which are more adequate than other methodologies.

## 3 Method

### 3.1 Data Collection

Following a thorough comparative analysis, the "DAVIS Driving Dataset 2020" (DDD20) was selected as the primary reference for this research [33]. Compared to its predecessor, DDD17, DDD20 offers a significantly larger volume of data, with 51 h of recordings over 4000 km of driving. The dataset includes detailed vehicle diagnostics collected via the OBDII port, read using the OpenXC interface. The available parameters include:

- Vehicle speed
- Steering angle
- Throttle position
- Brake status
- Other critical parameters for driving behavior modeling

Unlike other examined datasets, DDD20 also provides location information for some of its data, enabling the correlation between driving style and environmental context. Additionally, it includes recordings in various road scenarios (highways and urban roads) and under different weather conditions, enhancing the generalizability of the results.

Adopting the DDD20 dataset enables a comprehensive and methodologically robust analysis of driving behavior, leveraging reliable and contextually relevant OBD data. This selection ensures the consistency of the results and their applicability to real-world scenarios, providing a solid foundation for driving style modeling.

The other datasets studied but not considered to achieve the objective of this study are listed and briefly described in the several bullet points below:

- **HDD (Honda Research Institute Driving Dataset)**: Designed for training autonomous driving algorithms in urban environments, this dataset includes CAN-BUS data and video recordings of the surrounding context. However, since the data was collected with drivers strictly adhering to traffic regulations, it lacks significant variations in driving behavior and is therefore unsuitable for analyzing real-world driver behavior.
- **nuScenes**: Developed for testing autonomous systems, this dataset combines data from cameras, LIDAR, radar, and inertial measurement units. It includes CAN bus data similar to OBD information, such as vehicle speed, steering angle, and brake status. However, it lacks reliable GPS coordinates, preventing the analysis of the relationship between driving style and road context, a key requirement for this study.

- **Drive360**: A dataset focused on training autonomous driving models, it contains CANBUS data and navigation information. However, the available OBD data is limited to steering angle and speed, lacking critical parameters such as braking, acceleration, and engine-related variables, making it unsuitable for driving style classification.
- **DDD17**: A dataset integrating dynamic vision sensors (DVS) and OBD data to monitor vehicle behavior. While it includes useful information on speed and engine RPM, it does not provide location data, making it inadequate for a contextual analysis of driving behavior.
- **PoliDriving**: Developed for road safety research, this dataset includes OBD and location data. However, the provided GPS coordinates do not correspond to the actual data collection locations, and the dataset authors were unable to provide a conversion method, rendering it unusable for our study.

As it is possible to notice, the DDD20 dataset provides data from OBD devices, which is equipment that could be implemented and used easily and with lower costs than other technologies (i.e., LIDAR). Another advantage of the dataset, in comparison with others, is that DDD20 provides information on braking, accelerations, engine-related variables, and GPS coordinates (essential to feature context analysis).

### 3.2 Feature Selection

The identification of relevant characteristics for driving behavior analysis requires careful evaluation of the parameters available from the On-Board Diagnostics (OBD) system. This study selected 11 principal features to characterize the driver's driving style through a 15-s time window (see Table 1).

**Table 1.** Main available features with a brief description from the On-Board Diagnostics (OBD).

| Feature | Description |
|---|---|
| Mean accelerator pedal position | The average percentage of accelerator pedal pressure during the time window, indicating the driver's acceleration intent |
| Mean engine speed | Average engine rotations per minute (RPM), reflecting the engine's operational regime |
| Mean vehicle speed | The average velocity of the vehicle, providing insights into driving dynamics |
| Mean torque at the transmission | Average torque delivered to the transmission system, measuring the force applied for vehicle movement |
| Number of gear position changes | Count of transmission gear changes, indicating driving style aggressiveness |
| Brake pedal status variations | Number of brake pedal state changes, representing braking pattern behavior |
| Fuel consumption difference | Variation in fuel consumption during the time window, indicating efficiency |
| The sum of accelerator position variations | Cumulative changes in accelerator pedal position, measuring throttle control smoothness |

(*continued*)

**Table 1.** (*continued*)

| Feature | Description |
|---|---|
| The sum of vehicle speed variations | Cumulative changes in vehicle speed, indicating speed control consistency |
| The sum of engine speed variations | Cumulative changes in engine RPM, reflecting engine load management |
| The sum of steering wheel angle variations | Cumulative changes in steering angle, measuring steering behavior smoothness |

The selection of this time window was determined by analyzing the signal sampling frequency and the dynamics of the monitored features. This interval represents the minimum duration necessary to capture significant variations in vehicle characteristics, balancing temporal resolution with signal stability.

These characteristics were chosen for their ability to comprehensively represent driving dynamics and driver behavior. Each feature provides specific insights into different aspects of driving style, from basic vehicle control to efficiency patterns and aggressive behavior indicators.

### 3.3 Data Labeling

For driving style classification, an unsupervised labeling approach based on the K-means algorithm has been implemented [34, 35]. This method was chosen for its ability to identify natural behavioral patterns in the data without requiring predefined labels.

The labeling process consists of two main phases. In the first phase, K-means clustering with $k = 3$ is applied to identify three principal behavioral classes: normal driving, dangerous driving, and safe driving (see Table 2). This preliminary classification enables the grouping of similar driving behaviors based on features extracted from OBD data.

**Table 2.** Principal Behavioural Classes used for the labeling approach based on the K-means algorithm.

| Cluster | Description | Count |
|---|---|---|
| 0 | Safe Driving | 79,788 |
| 1 | Normal Driving | 25,027 |
| 2 | Dangerous Driving | 46,226 |

The second phase involves evaluating feature impact through a ranking system that assigns scores from 1 to 3 based on the comparison of feature means across clusters. This approach allows quantification of each feature's contribution in defining different driving styles and validates the consistency of the obtained classification.

The final ranking of each cluster is determined by summing the scores assigned to individual characteristics, enabling objective identification of the cluster associated with the most aggressive driving style, characterized by the highest overall score.

### 3.4 Geographical Data Enrichment

Following the labeling phase, the dataset is enriched with contextual information related to the road environment through integration with OpenStreetMap (OSM). This geographical enrichment process enables correlation between identified driving behaviors and characteristics of road infrastructure and urban context.

Using GPS coordinates associated with each acquisition time window, information is extracted from OSM including road type, speed limits, presence of intersections or roundabouts, and proximity to significant points of interest such as schools or commercial areas. This additional layer of information allows contextualization of identified driving styles relative to the surrounding environment, providing a deeper understanding of interactions between driver behavior and road infrastructure characteristics.

The integration of this territorial data is particularly relevant for road safety analysis, enabling the identification of potential correlations between specific driving patterns and urban context characteristics.

The data preprocessing encompasses two main phases. The first phase utilizes One-Hot Encoding [36] to transform categorical variables (such as weather conditions, road type, and points of interest) into a numerical format comprehensible for Machine Learning algorithms. The second phase applies MinMaxScaler [37] to normalize all features within the 0–1 range, ensuring a fair comparison between variables with different scales.

This procedure enables the preparation of both onboard sensor (OBD) data and environmental context data for model input, ensuring optimal processing while preserving the relationships between variables. This standardized approach is crucial for the subsequent classification model to accurately interpret the diverse range of input features.

## 4 Results

### 4.1 Classification Task

The Feed-forward Neural Network implemented is trained and tested by using OBD enriched with geographical information. The three driving classes identified are “Safe”, “Moderate”, and “Dangerous”.

The multi-layer classifier, based on a feed-forward neural structure, demonstrates its learning capabilities through detailed performance metrics analysis during the training and validation phases. The experimental results show an efficient and robust optimization process, characterized by well-defined convergence patterns.

The loss function, as shown in Fig. 1, exhibits a rapid initial decrease, dropping from 0.30 to approximately 0.02 within the first 10 epochs, subsequently stabilizing. Concurrently, the validation accuracy steadily improves, reaching and maintaining values above 0.98 after the initial training phases (see Fig. 2).

The learning curves, illustrated in Fig. 3, demonstrate balanced performance evolution across both the training set and cross-validation (three-fold cross-validation). The neural network achieves accuracy values exceeding 0.98, with the training curve slightly above the validation curve, indicating an optimal balance between bias and variance, without evident signs of overfitting or underfitting.

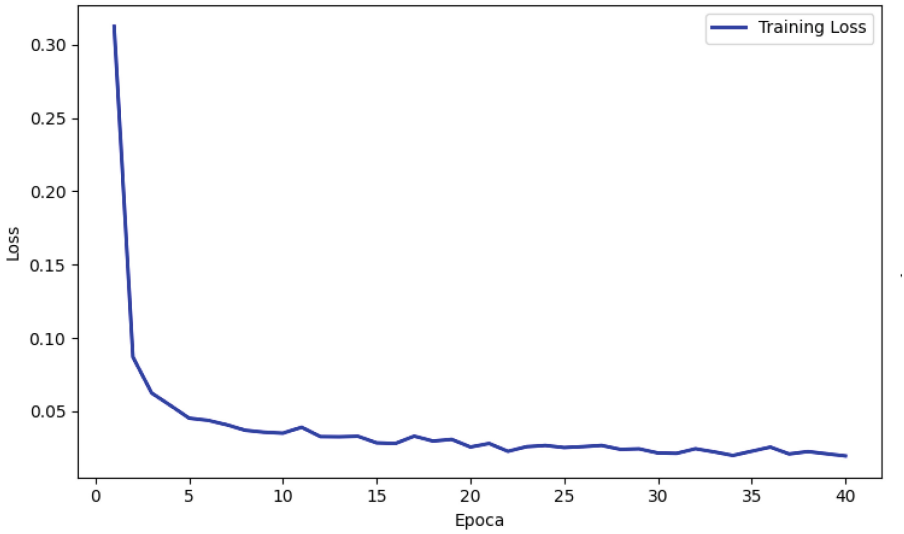


**Fig. 1.** Loss function during the training phase through different epochs.

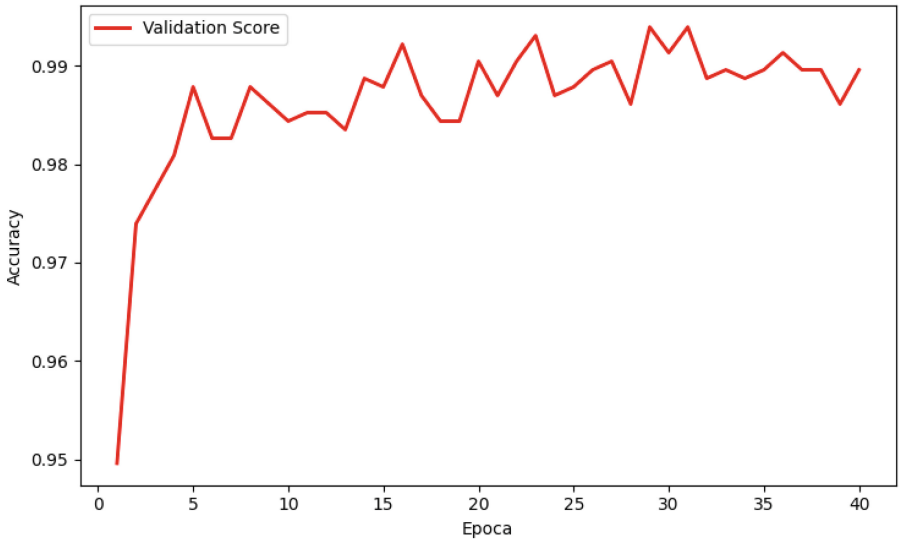


**Fig. 2.** Accuracy trend during the training phase through different epochs.

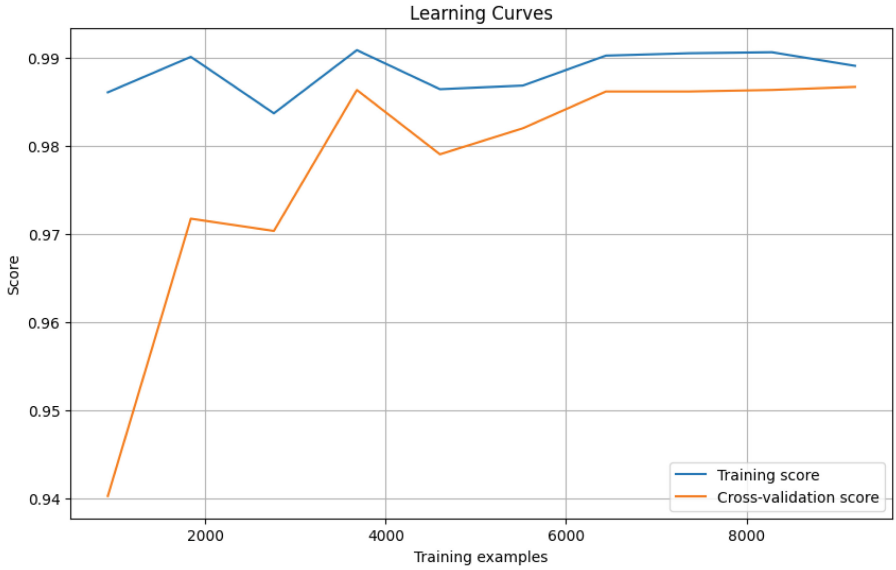


**Fig. 3.** Graph of the accuracy metric obtained for the model during training with different dataset sizes to conduct a cross-validation study.

### 4.2 Classification Results

The classification performance analysis, visualized through the confusion matrix (see Figs. 4 and 5), highlights the system's high accuracy in discriminating between three driving style classes. The network correctly classified 224 cases as "dangerous", 2615 as "moderate", and 2040 as "safe". Classification errors remain minimal and primarily confined to adjacent classes: 7 "dangerous" cases misclassified as "moderate", 19 "moderate" cases misclassified as "safe", and 23 "safe" cases misclassified as "moderate". Performance metrics show accuracies of 97.0% for the "dangerous" class, 99.2% for "moderate", and 98.9% for "safe", confirming the robustness and reliability of the approach in identifying different driving styles.

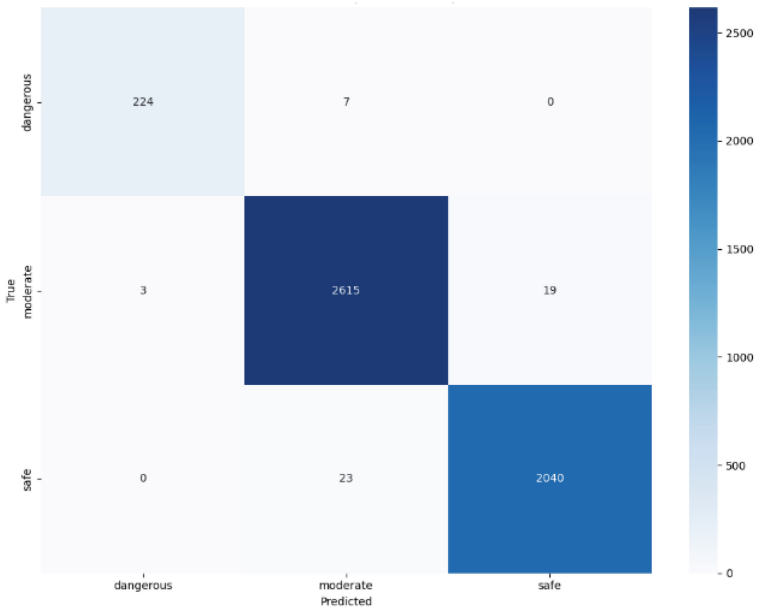


**Fig. 4.** Confusion matrix (absolute values) of the driving style classification model using a combination of OBD data and geographic context data.

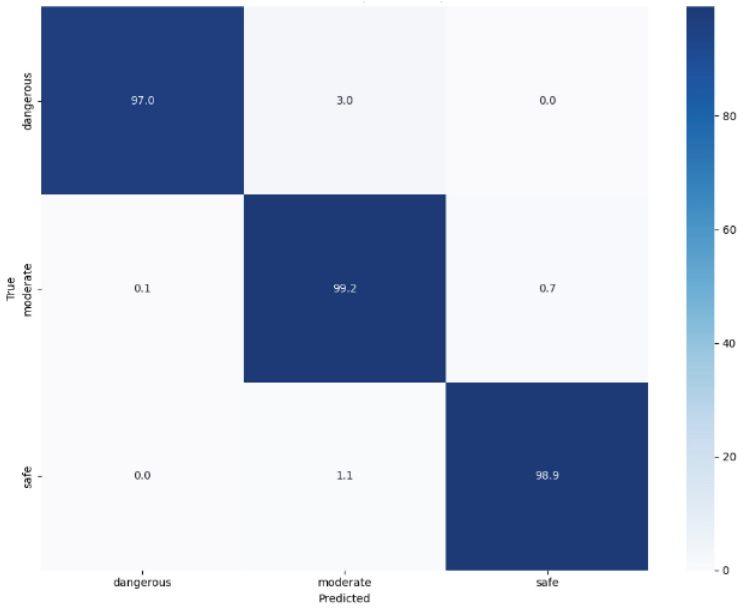


**Fig. 5.** Confusion matrix (percentages) of the driving style classification model using a combination of OBD data and geographic context data.

## 5 Conclusions

The introduction and the subsequent spread of information and communication technology in road vehicles (ICT), alongside the implementation of machine learning algorithms, has enabled a better understanding of driver behaviors. In vehicle data systems, smartphones, and real-time behavior detection methods are often implemented to collect the data required: Neural Networks, Support Vector Machines, and Decision Trees, are often used to process this data.

Human errors are the leading cause of road accidents and according to the literature, eco-driving behaviors can reduce both fuel consumption and unsafe driving. There is a strong use in the literature of studies that utilize OBD systems and machine learning to classify driving behaviors and predict fuel consumption.

In this paper, the first step was the collection of data. After the analysis of several datasets, the most adequate has resulted in the DDD20 dataset. This contains a larger amount of data than the other ones, with 51 h and 4000 km of total driving times and distances. To reduce the number of features contained in the datasets and to avoid overfitting, feature selection procedures and algorithms are implemented. An Unsupervised labeling approach based on K-means is also used for data labeling. To enrich the dataset, geographical data are added to the analysis to take into account the characteristics of infrastructures and urban contexts to predict driver behaviors. The findings show good prediction performances of the Feed-forward Neural Network in terms of accuracy when OBD and geographical data are both used.

The main contribution of this paper in the driver behavior field is the use of geographical data to enrich analysis with infrastructure and urban-related features, to classify safe and unsafe driver behaviors more accurately.

One of the limitations of this study relies on the absence of feature importance analysis to highlight the contribution of infrastructure context-related variables to driver behavior, but this is also a natural future development of this work. So, the next step in this research is to implement Variable Importance Analysis (such as SHAP, SHapley Additive exPlanations) to make this model more interpretable. A limitation of the study, despite the high accuracy of the models implemented, lies in the lack of interpretability. It is difficult to understand which is the role that the single feature plays in the model output. As further development, the challenge is to try to predict driver behaviour by using only geographical information that could be retrieved easily, without the implementation of devices and specific equipment to collect data on board.

The high performance obtained by combining OBD data and geographical context information shows that context features play a predominant role in predicting driver behaviours. This also highlights that this model could be adapted to find potential critical areas or zones in which it is essential to improve safety conditions, but also environmental conditions. In this perspective, road managers and administration could be provided with a useful and simple tool to promote new policies and prioritize interventions.

**Acknowledgments.** The research reported in this paper was developed in the EMOTIVEs Research Project, supported by the "FESR Lazio 2021-2027 Program—RSI Competitive Repositioning Notice".

**Disclosure of Interests.** 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. The authors have no competing interests to declare that are relevant to the content of this article.

## References


1. Peppes, N., Alexakis, T., Adamopoulou, E., Demestichas, K.: Driving behaviour analysis using machine and deep learning methods for continuous streams of vehicular data. Sensors. **21**(14), 4704 (2021)
2. Pholprasit, T., Choochaiwattana, W., Saiprasert, C.: A comparison of driving behaviour prediction algorithm using multi-sensory data on a smartphone. In: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), pp. 1–6. IEEE (2015, June)
3. Chan, T.K., Chin, C.S., Chen, H., Zhong, X.: A comprehensive review of driver behavior analysis utilizing smartphones. IEEE Trans. Intell. Transp. Syst. **21**(10), 4444–4475 (2019)
4. Al-Hussein, W.A., Kiah, M.L.M., Yee, L., Zaidan, B.B.: A systematic review on sensor-based driver behaviour studies: coherent taxonomy, motivations, challenges, recommendations, substantial analysis and future directions. PeerJ Comput. Sci. **7**, e632 (2021)
5. Alluhaibi, S.K., Al-Din, M.S.N., Moyaid, A.: Driver behavior detection techniques: a survey. Int. J. Appl. Eng. Res. **13**, 8856–8861 (2018)
6. Arifoglu, D., Bouchachia, A.: Activity recognition and abnormal behaviour detection with recurrent neural networks. Procedia Comput. Sci. **110**, 86–93 (2017)
7. Abou Elassad, Z.E., Mousannif, H., Al Moatassime, H., Karkouch, A.: The application of machine learning techniques for driving behavior analysis: a conceptual framework and a systematic literature review. Eng. Appl. Artif. Intell. **87**, 103312 (2020)
8. Larsson, P., Tingvall, C.: The safe system approach—a road safety strategy based on human factors principles. In: Engineering Psychology and Cognitive Ergonomics. Applications and Services: 10th International Conference, EPCE 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21–26, 2013, Proceedings, Part II 10, pp. 19–28. Springer, Berlin/Heidelberg (2013)
9. U.S. Department of Transportation: What is a Safe System Approach? https://www.transportation.gov/NRSS/SafeSystem. Last Accessed Dec 2024
10. Petridou, E., Moustaki, M.: Human factors in the causation of road traffic crashes. Eur. J. Epidemiol. **16**, 819–826 (2000)
11. Ellison, A.B., Greaves, S.P., Bliemer, M.C.: Driver behaviour profiles for road safety analysis. Accid. Anal. Prev. **76**, 118–132 (2015)
12. Somayajula, D., Meintz, A., Ferdowsi, M.: Designing efficient hybrid electric vehicles. IEEE Veh. Technol. Mag. **4**, 65–72 (2009)
13. Yen, M.H., Tian, S.L., Lin, Y.T., Yang, C.W., Chen, C.C.: Combining a universal OBD-II module with deep learning to develop an eco-driving analysis system. Appl. Sci. **11**(10), 4481 (2021)
14. Xu, N., Li, X., Liu, Q., Zhao, D.: An overview of eco-driving theory, capability evaluation, and training applications. Sensors. **21**(19), 6547 (2021)
15. Graba, M., Mamala, J., Bieniek, A., Sroka, Z.: Impact of the acceleration intensity of a passenger car in a road test on energy consumption. Energy. **226**, 120429 (2021)
16. Loman, M., Šarkan, B., Skrúcaný, T.: Comparison of fuel consumption of a passenger car depending on the driving style of the driver. Transp. Res. Procedia. **55**, 458–465 (2021)

17. Shahverdy, M., Fathy, M., Berangi, R., Sabokrou, M.: Driver behavior detection and classification using deep convolutional neural networks. Expert Syst. Appl. **149**, 113240 (2020)
18. Kaplan, S., Guvensan, M.A., Yavuz, A.G., Karalurt, Y.: Driver behavior analysis for safe driving: a survey. IEEE Trans. Intell. Transp. Syst. **16**(6), 3017–3032 (2015)
19. Kang, H.B.: Various approaches for driver and driving behavior monitoring: a review. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 616–623 (2013)
20. Pratama, B.G., Ardiyanto, I., Adji, T.B.: A review on driver drowsiness based on image, bio-signal, and driver behavior. In: 2017 3rd International Conference on Science and Technology-Computer (ICST), pp. 70–75. IEEE (2017, July)
21. Pugeault, N., Bowden, R.: Learning pre-attentive driving behaviour from holistic visual features. In: Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part VI 11, pp. 154–167. Springer, Berlin/Heidelberg (2010)
22. Cyganek, B., Gruszczyński, S.: Hybrid computer vision system for drivers' eye recognition and fatigue monitoring. Neurocomputing. **126**, 78–94 (2014)
23. Lattanzi, E., Castellucci, G., Freschi, V.: Improving machine learning identification of unsafe driver behavior by means of sensor fusion. Appl. Sci. **10**(18), 6417 (2020)
24. Fugiglando, U., Massaro, E., Santi, P., Milardo, S., Abida, K., Stahlmann, R., Netter, F., Ratti, C.: Driving behavior analysis through CAN bus data in an uncontrolled environment. IEEE Trans. Intell. Transp. Syst. **20**, 737–748 (2019)
25. Fadzil, N.F.A.M., Fadzir, H.M., Mansor, H., Rahardja, U.: Driver behaviour classification: a research using OBD-II data and machine learning. J. Adv. Res. Appl. Sci. Eng. Technol., 51–61 (2024)
26. Uvarov, K., Ponomarev, A.: Driver identification with OBD-II public data. In: 2021 28th Conference of Open Innovations Association (FRUCT), pp. 495–501. IEEE (2021, January)
27. Ameen, H.A., Mahamad, A.K., Saon, S., Malik, R.Q., Kareem, Z.H., Bin Ahmadon, M.A., Yamaguchi, S.: Identification of driving safety profiles in vehicle to vehicle communication system based on vehicle OBD information. Information. **12**(5), 194 (2021)
28. Abukhalil, T., AlMahafzah, H., Alksasbeh, M., Alqaralleh, B.A.: Fuel consumption using OBD-II and support vector machine model. J. Robot. **2020**(1), 9450178 (2020)
29. Rivera-Campoverde, N.D., Muñoz-Sanz, J.L., Arenas-Ramirez, B.D.V.: Estimation of pollutant emissions in real driving conditions based on data from OBD and machine learning. Sensors. **21**(19), 6344 (2021)
30. Lattanzi, E., Freschi, V.: Machine learning techniques to identify unsafe driving behavior by means of in-vehicle sensor data. Expert Syst. Appl. **176**, 114818 (2021)
31. Kumar, R., Jain, A.: Driving behavior analysis and classification by vehicle OBD data using machine learning. J. Supercomput. **79**(16), 18800–18819 (2023)
32. Abdelrahman, A.E., Hassanein, H.S., Abu-Ali, N.: Robust data-driven framework for driver behavior profiling using supervised machine learning. IEEE Trans. Intell. Transp. Syst. **23**(4), 3336–3350 (2020)
33. DAVIS Driving Dataset 2020 (DDD20): http://sensors.ini.uzh.ch/databases.html. Last Accessed 11 Feb 2025
34. Coates, A., Ng, A.Y.: Learning feature representations with k-means. In: Neural Networks: Tricks of the Trade, 2nd edn, pp. 561–580. Springer, Berlin/Heidelberg (2012)
35. Bouhsissin, S., Sael, N., Benabbou, F.: Driver behavior classification: a systematic literature review. IEEE Access. **11**, 14128–14153 (2023)

36. Okada, S., Ohzeki, M., Taguchi, S.: Efficient partition of integer optimization problems with one-hot encoding. Sci. Rep. **9**(1), 13036 (2019)
37. Brownlee, J.: How to Use StandardScaler and MinMaxScaler Transforms in Python (2020). https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python (2021)