Eco-friendly Naturalistic Vehicular Sensing and Driving Behaviour Profiling

Doctoral Dissertation
Doctor of Philosophy
Interactive and Cognitive Environments (JD ICE) course
November 2019

Rana MASSOUD

UNIVERSITY OF GENOA (UNIGE)

UNIVERSITÀ DEGLI STUDI
DI GENOVA

&

QUEEN MARY UNIVERSITY OF LONDON (QMUL)

Queen Mary
University of London

A dissertation submitted to the Dipartimento di Ingegneria Navale, Elettrica, Elettronica e delle Telecomunicazioni (DITEN), Elios Lab (UNIGE) and School of Electronics Engineering and Computer Science (EECS), IoT2US Lab (QMUL) in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Electronic Engineering and Computer Science

Dissertation committee

Supervisors:

Prof. Francesco BELLOTTI (UNIGE University of Genoa, Italy)
Prof. Stefan POSLAD (Queen Mary University of London, UK)

Independent assessors:

Prof Mehrdad Dianati (University of Warwick, UK)
Dr. Pratheep Kumar Paranthaman (Elon University, USA)

Abstract

Internet of Things (IoT) technologies are spurring of serious games that support training directly in the field. This PhD implements field user performance evaluators usable in reality-enhanced serious games (RESGs) for promoting fuel-efficient driving. This work proposes two modules – that have been implemented by processing information related to fuel-efficient driving – to be employed as real-time virtual sensors in RESGS. The first module estimates and assesses instantly fuel consumption, where I compared the performance of three configured machine learning algorithms, support vector regression, random forest and artificial neural networks. The experiments show that the algorithms have similar performance and random forest slightly outperforms the others. The second module provides instant recommendations using fuzzy logic when inefficient driving patterns are detected. For the game design, I resorted to the on-board diagnostics II standard interface to diagnostic circulating information on vehicular buses for a wide diffusion of a game, avoiding sticking to manufacturer proprietary solutions. The approach has been implemented and tested with data from the enviroCar server site. The data is not calibrated for a specific car model and is recorded in different driving environments, which made the work challenging and robust for real-world conditions. The proposed approach to virtual sensor design is general and thus applicable to various application domains other than fuel-efficient driving. An important word of caution concerns users' privacy, as the modules rely on sensitive data, and provide information that by no means should be misused.

Acknowledgements

I acknowledge that this research was partially funded as part of a Joint Doctorate Interactive and Cognitive Environments (JD-ICE) between University of Genova (UNIGE) Dipartimento di Ingegneria Navale, Elettrica, Elettronica e delle Telecomunicazioni, Elios Lab and Queen Mary University of London (QMUL), School of Electronic Engineering and Computer Science, IoT2US Lab.

Indeed, I feel in deep gratitude to my UNIGE supervisor Prof Francesco Bellotti, and my QMUL supervisor Prof Stefan Poslad, for their (direct and remotely) continuous follow-up, guidance, suggestions, advice. I would like to thank Dr Stefan for coming to UNIGE towards a direct project meeting, at the beginning of this research. In short, I learn a lot from their immense knowledge and I appreciate their kind support, patience, motivation, and enthusiasm.

Thanks to my family (especially my mother) and my friends for their support and encouragements during my three years PhD, so that I am now where I stand.

I also owe a great debt of gratitude to the academic board of the course especially Prof. Carlo Regazzoni. I sincerely thank also Prof Mehrdad Dianati and Dr Pratheep Kumar Paranthamanand for being external reviewers.

I acknowledge technical support given by enviroCar open Citizen Science platform for traffic monitoring and environment (in particular Christoph Stasch and Arne de Wall), led by several project partners, including 52north with whom I liaised.

Table of Content

Table of Figures

List of Tables

Acronyms

CASES

1 Introduction

In addition to smarter road vehicles and smarter roads, more smartly human-driven vehicles (that are as yet still predominantly more common-place than autonomously driven ones) currently still have a huge potential to improve road safety, fuel-efficiency and to reduce vehicle exhaust emissions. Reality-enhanced serious games (RESGs) are emerging serious games (SGs) genre – games for training and learning purpose beyond behavioural improvement via users' motivation and engagement in a digital environment – that has the ability to contextualize a game within its real instruction-target environment. This is by incorporating data from the real world to enact training in the wild. A key module for such games is the evaluator, that senses a user performance and provides consequent input to the game. This – with the proper cautions due to safety - can be done also for daily activities, such as driving.

The remaining of this chapter is organised as it follows, section 1.1 introduces the importance of the problem of interest that this research faces; section 1.2 presents the challenges in the estimation of fuel consumption (FC) and the difficulty of driving style evaluation based on the achieved fuel efficiency; section 1.3 reviews quickly the previous driving study domains; section 1.4 defines the research objective; while section 1.5 overviews the rest of this thesis.

1.1 Motivation: Improving Vehicle Driver Driving Behaviour with Gamification Motivation towards Fuel Saving

1.1.1 The problem of interest: Inefficient Driving Styles can Diminish Fuel Economy

Inappropriate (aggressive) driving styles increase fuel wastage especially for high loaded vehicles (e.g., heavy-duty trucks), cause safety predicament, wastes a lot of energy and deteriorate engine health [1]. The difference in FC between a normal and an aggressive driving pattern is estimated to be above 40%, in favor of the normal driving style [2]. Although given the selection of a vehicle (purchasing fuel-efficient vehicles) can have the most effect on fuel economy, inefficient driving style besides lack of following eco-driving general advice (e.g., lack of attention to maintenance practices, route selection, and managing vehicle load), can diminish fuel economy up to 45% [3].

1.1.2 Problem Importance: Fuel Saving Potential via Gamification' Motivation

Improving vehicle' driver behaviour still has a significant potential to increase fuel-efficiency and road safety, and reduce emissions in addition to smarter road vehicles and smarter roads [4][5][6]. It has been estimated that vehicle drivers can save up to 25% of fuel by adopting appropriate (efficient) driving patterns with simple practices (e.g., optimal clutch and engine rotation) [7][8][9] with variations depending on the type of a vehicle (it could be less important in modern vehicles) [10].

Motivation can help drivers to drive more efficiently [15][16]. It has been shown that eco-driving advice supports fuel-saving [12], where providing continuous feedback in the design of eco-driving assistants is recommended to improve the effects achieved by learning [17][18]. Therefore, there is a need to provide drivers with proper and understandable advice. Such eco-driving advice can decrease fuel consumption (FC) from 5 to 25% [9].

Vehicles' drivers have significant margins to improve safety and reduce emissions (1.1.1) and this could be promoted via gamification. Gamification appeals to the human instinct to compete. It leverages people's natural desires for status, achievement, competition and to be part of an inclusive social community. A well-designed SG can be a significant instrument towards eco-driving motivation [15]. A SG can improve the user experience by combining training and entertainment. SG techniques have been trialled in the automotive and transportation sector because of their motivating and inspiring potential [19][20]. Moreover, the emerging genre of RESGs, which stems also from the spreading of the Internet of Things (IoT) technologies, seems particularly relevant in this regard.

In the emerging genre of reality-enhanced serious games (RESGs), in-game progress is due not only to the digital gaming ability of the player, but also the game inputs are provided particularly by real-world measurements by sensing a user's performance in the actual target field [21][22]. This is an evolution of pervasive gaming [24], where the game's fictive world blends with the physical world connecting a digital game environment with reality, and allows opening and exploiting a direct, possibly real-time (RT), link between a game and a training objective. Therefore, field users' performance becomes a key factor [22] and should be easily understandable to supply effective coaching feedback to players. By using RT driving data, this technology can impact drivers' attitude towards more economic driving and greater safety. Also delivering this RT driving data to the administrators in fleet management domain – who are keen on reviewing their drivers' driving patterns – can help achieve company-wide goals, productivity and compliance.

1.2 Challenges

As an indicator of driver performance in this study, I relied on FC, which is strongly influenced by driving styles [6][2][7][9][25] and can be quantified and validated.

1.2.1 The Need to Estimate Fuel Consumption

FC' modelling is of prime importance with the constantly rising price of fuel, as well as due to ecological reasons. FC is not directly accessible through the common On-Board Diagnostic-II (OBD-II) interface, as the "engine fuel rate" is only readable in rare car models (it is not mandatory in the OBD-II standard protocol) [26]. Therefore, it has to be estimated from other available OBD-II standard signals (supported to all cars) and information.

In general, the estimation of FC is challenging, taking the fact that it is influenced by several factors apart from driving patterns (e.g., driving environment, vehicle maintenance) [27][28]. This corresponds to the general case in which a user's performance should be estimated by considering several different influencers.

1.2.2 The difficulty of Evaluating Driving Style

The recognition of driving style is a vague notion due to the difficulty of objectively evaluating human driver performance. Driving capabilities depend on drivers' age, driving experience, his/her physical, emotional and psychological/mental state.

Yet, there is no unique definition of driving behaviour, nor a unique standard measure for its evaluation since it is a combination of mixed factors and components [14]. Also, checking whether a driver is adopting an eco-driving style, is not an easy task, since the eco-driving approach is a wide concept as discussed later in 2.2.

1.2.3 The complexity of Evaluating Driving Pattern Based on the Fuel Consumption

Fuel economy for the same car may vary between several drivers depending on their habits [11]. A careless person uses to drive with aggressive accelerating and deceleration attitude. A calm driver acts more consciously with the accelerator and the brake pedals for more fuel economy.

However, other external and internal influences result in consuming more fuel apart from driving style (Fig. 1). Starting with the vehicle's performance factor. Purchasing fuel-efficient vehicles can have the most effect on fuel economy – the higher the values of the number of cylinders, number of engine displacements (number of CC), number of horsepowers, car weight and engine size, the more a car consumes fuel. Adding to that, fuel economy is impacted by car cost. Vehicles become more efficient with the improvement of in-vehicle technologies with manufacturers. Also, the state of vehicle maintenance impacts fuel-efficiency (e.g., tire not properly inflated). Another factor is the engine's load of a vehicle, which is affected by weight carried by the vehicle such as the number of passengers, or because of car configuration (e.g., heated seats, headlights, entertainment equipment).

Driving environment – where some of them are hard to quantify – has a great impact on driving style, thus affects fuel economy, such as driving condition (e.g., traffic jam in the city, traffic light timings), road types (such as a highway), road infrastructure (e.g., potholes, bends, number of intersections and lanes), daytime periods (e.g., rush hours in the morning and afternoon period, weekday or weekend days). Adding to this, weather conditions (precipitation, humidity, icy roads, etc.) influence applicable speed profiles – a critical input for driving analytics since it is positively strongly correlated with FC. For instance, rainy days cause traffic (the cars to be in idling state – vehicle speed is null while RPM is not null) and then more fuel is consumed. Also, in hot days (above 30 ◦C), drivers turn on the air conditioning leading to burning more fuel.

1.3 Summary of the Previous Driving Study domains Solutions

Drivers' behaviours have been studied through years for different vehicle applications, where improving vehicle's fuel-efficiency has gained special attention in the majority of those studies. Driving behaviour has been one of the five behaviours related to personal vehicular transportation for reducing household sector emissions (purchase of fuel-efficient vehicles, low rolling resistance tires, routine auto maintenance, driving behaviour and carpooling) [23].

Thus, over the last few years, interest has been raised in monitoring vehicles and driving data in different application contexts, aiming to identify driving situations and manoeuvres that are risky, reduce the energy efficiency and increase engine emissions. Also, promoting energy-efficient driving with eco-driving approach has received scarce attention from the research community [11]. For example, the administrators of fleet management domain are keen on reviewing their drivers' driving styles through gathering more finely-grained information about fleet usage, which is influenced by driving patterns.

Also, car insurance firms could place insurance premiums (additional costs) based on someone's past driving records with Pay-As-You-Drive (PAYD) or Pay-How-You-Drive (PHDR) or Pay-As-You-Go (PAUG) strategies [29]. Such mile-based insurance famous providers (e.g., Progressive, Allstate and Geico) – applauded by environmentalist groups – believe that the combination of context, awareness and financial incentive work to convince people to drive less. Their strategies consist of combining a fixed monthly fee with a variable fee based on the distance travelled. This encourages then drivers to reduce the amount of time on the road or restrict what times they are driving for saving money. These options lower the risk of offering new drivers covers from car insurance companies. In return, drivers who do not drive as much as commuters and are often unable to pay high rates (e.g., students), can pay cheaper premiums while they build up their no-claims [30].

Furthermore, more recently governments may require drivers to drive more efficiently for receiving a driving license, e.g., in Spain [10]. On the other hand, characterizing human driving pattern is fundamental in the application of modern transportation services such as the application of artificial intelligence (AI) in the transport sector, and the connected autonomous vehicles for safer (e.g., reduce the human errors), cleaner, smarter and more efficient transport modes [11][13][14]. A review of previous related work is presented later in "2 Background and Survey".

Fig. 1. Factors that affect fuel consumption.

1.4 Research Objectives: Contribution and Expected Impact

This work targets to aggregate feedback on drivers' driving and creates a driving specific profile in order to improve their skills for adopting eco-friendlier driving styles. It proposes a methodology through which to process vehicular data in RT, so to compute driver performance assessment values (specifically, on driving efficiency – the less the fuel is consumed, the more the driving is efficient). The study analyses naturalistic vehicular driving data for producing information related to eco-driving (e.g., achieved fuel efficiency), and driving pattern based on information extracted from on-board in-car sensors (e.g., engine speed and car speed) using On-Board Diagnostics-II system and smartphones embedded sensors via OBD-II Bluetooth/WiFi/wired adapters.

The proposed approach may be integrated by game designers as field user performance evaluators in third-party RESGs to assist automotive' drivers in promoting more fuel-efficient driving (low FC) as vehicles become ever more powerful IoT platforms [31] and regarding the positive role of eco-driving motivation on driving style (discussed later in 2.2) that can be achieved via gamification motivation [4](discussed later in 2.3.4). The presented gaming approach for driving profiling might be beneficial also for insurance firms, fleet management and driving schools.

There are already automakers who relied on gamification motivation such as Honda's Eco Assist which rewards 'green' driving styles by the number of leaves displayed in the instrument cluster to indicate the level of driving performance from the environmental side (discussed later in 2.3). To the best of my knowledge, the technological details of the manufacturer-specific solutions towards efficient drivers are not disclosed in the literature. Following this industrial direction, I intend to support the development of new gameful solutions for fuel efficiency improvement, through a tool available to all the drivers, by using the standard OBD-II vehicular interface and exploiting machine learning (ML) and AI algorithms to extract RT information about a drive and to support coaching.

Data from enviroCar project's, a naturalistic driving archive which freely provides a significant amount of naturalistic drive data recorded in numerous vehicles and driving conditions (described in section 3), was used in the analysis, implementation and the simulation. This data includes estimated FC (with a mathematical formula eq (2) described in 3.2 (4)) that it could be used to develop a new model based on machine learning (ML).

User assessment information should be easily accessible by games' designers. I thus resorted to the on-board diagnostic (OBD-II) standard interface [35] (described later in 2.1.2) to the diagnostic information circulating on vehicular buses as a fundamental asset for an automotive game context. This should allow developers to create games for every kind of vehicles avoiding sticking to the original equipment manufacturer (OEM) proprietary solutions.

The outcomes of this research have been documented in four accepted/published scientific articles (one journal and three publications in international conferences). Furthermore, an article was submitted to the "IEEE Transactions on Games" in December 2019. Those publications which are listed at the end of this report (before the "Bibliography"), illustrate my PhD research activities (described in section 4).

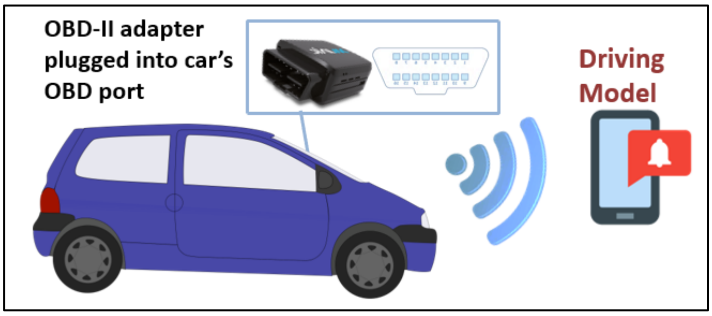


Fig. 2. Methodology of the proposed approach.

The target of this thesis could be summarized in two main objectives; (i) instant assessment with FC estimation values through a proper game mechanic (e.g., the energy available to the player avatar in a PassengerGame in [32]) detailed in 1.4.1 and (ii) RT recommendations while driving as presented in 1.4.2. Both sub-modules for implementing both objectives, can be employed as virtual sensors for driver's behaviour assessment via a RESG implemented on a smartphone (or on a raspberry pi) connected via Bluetooth to the OBD-II adapter as depicted in Fig. 2.

1.4.1 Objective 1: Instant Assessment with Fuel Consumption Estimation

Reality-enhanced gaming is an emerging SG genre, that can contextualize a game within its real instruction-target environment (described later in 2.3.3). A key module for such games is the evaluator, that senses a user performance and provides consequent input to the game. As stated in 1.2, the main indicator of driver performance in this study is the FC, which is strongly influenced by driving styles [2][9][25] and can be quantified and validated (1.1). FC measurements, however, are not directly accessible through the OBD-II interface (1.2.1). Thus, this research targets to estimate FC by requiring an elaboration of correlated signals. This reflects the general and common case in which user field performance is estimated by processing several information sources.

Drivers can be supplied with RT fuel-efficient fulfilment via a gameful design, such as the number of leaves (an interactive display integrated into the cockpit, with a plant metaphor) in Honda's Eco Assist (short for 'Ecological Drive Assist System') [33]. Those games' motivation elements are expected to encourage continuous improvement towards more fuel-efficient and safer driver pattern. Moreover, the continuous sequential determinations of FC can be realised directly by a game design as an ongoing updated score, or as the energy of the player, or to activate bonuses or maluses, or to facilitate reaching a higher gaming level, etc. (e.g., [22]).

Furthermore, it's beneficial to use gamification tools to score efficient behaviour of drivers after each trip such as on a scale of 100 (100 is the best possible score), and "gamify" it – through proper game mechanics in order to rank the performance of a driver against drivers' peers. This evaluating allows considering how dynamic driving behaviour is sometimes affected by other factors (e.g., traffic congestion). The more the driving becomes efficient, the higher the score (closer to 100) is earned by the driver. While the lower the score, the more the driving is inefficient and aggressive. Based on the final score, the driving pattern can be classified (such as fuel 'Saver') besides, a summary report for the driving pattern which could be useful.

1.4.2 Objective 2: Instant Driving Recommendation

An aggressive driving pattern (such as forced acceleration) results in consuming more fuel than a normal driving style. Given the effectiveness of monitoring a driver's behaviour (1.1) [34], I added a RT coaching functionality to the module. The module triggers direct driving advice considering what actions the drivers could do, to better control the fuel economy when more inefficient and riskier (aggressive) driving manoeuvres are detectable while driving.

The direct recommendation is based on sensing and analysing the changes of throttle position (TPS), engine revolutions per minute (RPM) and car speed. Yet, significant changes in those predictors are detected as FC-relevant events, e.g., signalling to overtake. Those vehicular signals are controllable by drivers and affect the fuel economy (proved in 3.5 and 4.1).

Moreover, relying on TPS, RPM and car speed is easy to explain when returning coaching feedback to the drivers. This is important for a wide diffusion of a game (the same model can be used by game' developer for different types of games) where users' performance should be easily understandable by game developer for supplying coaching feedback to the players: identify effectively different performance factors and perhaps manage them actively. This aims to improve the user' experience and strengthening the link with the actual activities to be improved by updating or adapting players' understanding and performance.

Such manoeuvres derived from the OBD-II interface to the vehicle sensors complemented by additional sensors such as GPS location, e.g., from a mobile phone or inbuilt vehicle sensor, can be regarded as eco-driving events to be provided such as eco-driving feedback via the user interface of a SG for keeping the driver aware of the fuel economy and safety.

The RT feedback improvement actions can be provided in the mean of voice prompts and/or other means suited to the driving environment, that lessens users' attention away from driving, else the use of SGs in the automotive domain may distract the driver, resulting in violations of traffic laws [36].

1.5 Outline

This report is organized into five main chapters as follows.

1. The first chapter "Introduction" covers the research's motivation (1.1), the difficulties of evaluating driving behaviour (1.2), 1.3 reviews quickly the previous driving study domains and it defines the research goal (1.4).
2. The second chapter "Background and Survey" covers the points; 2.1 introduces the manner to access internal vehicular data through automobile communication protocols; 2.2 defines the eco-driving concept, reviews its rules, its coaching studies and its training categories towards fuel saving. Furthermore, some controversial issues in the definition of this approach are discussed; 2.3 shows the significance of gamification's motivation towards keeping drivers aware of fuel-efficiency. 2.4 introduces the technologies that I have used for creating the modules.
3. The third chapter "Data Source, Acquisition, Manipulation and Primary Experiments" presents the driving data for the modelling and simulation stages of this study (3.1 and 3.2). Yet, it introduces the data acquisition process (3.3), data pre-processing (3.4) and some primary experiments (3.5).
4. The fourth chapter "Methodology" provides the followed approach towards achieving the project' target. It compromises four sub approaches in 4.1, 4.2, 4.3 and 4.4 respectively:
    a. Methodology 1 (4.1): I departed using the mathematical tool fuzzy logic (FL), given its ability to embody expert knowledge and to deal with incomplete availability of information. I relied on the key signals TPS, car speed and RPM.
    b. Methodology 2 (4.2): After the implementation of the FL model in 4.1, that can provide coaching advice to drivers besides the prediction of FC, I was interested in exploring if

better quantitative FC estimations (with the signals as in 4.1) could be obtained with random forest (RF) ML technique. The combination of both techniques can supply coaching advice to drivers via the deduced fuzzy rules (4.1), with a more accurate quantitative FC estimation that may be obtained through the RF model (that can be integrated into the game as an energy factor).

c. Methodology 3 (4.3): I work on a new approach for supplying the drivers with direct feedback when no eco-driving events are detectable considering what actions they could perform toward eco-friendlier attitude. Besides the already used three key signals (TPS, car speed and RPM), I also involved the estimated car jerks (changes in acceleration with respect to time). The thresholds for the sensors' evaluation have been defined based on the literature review. Again, the quantitative estimations of the FC with the RF module in 4.2, have been improved by involving the OBD-II calculated engine load predictor. On the other hand, considering that driving circumstances may lead to wasting more fuel (e.g., stop traffic light timings), it is beneficial to provide an eco-driving score on a scale of 100 for the trip, together with a summary report for the driving pattern. In this process, I involved the achieved fuel-efficiency (the most important metric in eco-driving).

d. Methodology 4 (4.4): I investigate three ML techniques, support vector machine for regression (SVR), RF and artificial neural networks (ANNs). I involved eleven predictors that affect the consumption of fuel, relying on vehicular signals available through the common ODB-II interface for a wide diffusion and a standardized FC module. This covers the case of unreadable or faulty MAF (3.2 (4)). I relied on the deduced fuzzy rules (in 4.1) for supplying the drivers with driving advice during driving to avoid fixing a threshold for each one of the inputs as in 4.3.

5. The conclusion, discussion and future work are drawn in the last chapter.

2 Background and Survey

This chapter reviews some crucial relevant concepts. It presents a survey for the related previous studies, findings and critical analyses to depart my study. 2.1 presents the manner to access internal vehicular data through automobile communication protocols. Section 2.2 defines the eco-driving term, besides a summary of its rules, its coaching studies and its training categories towards keeping drivers aware of fuel economy. Furthermore, it discusses some controversial issues in the definition of this approach. The significance of gamification's motivation towards keeping drivers aware of fuel-efficiency is summarized in 2.3. Section 2.4 anticipates the adopted techniques in these research methodologies that are detailed later in section 4.

2.1 Automobile Sensing and Vehicular Data Communication Protocols

Driving data analysis is ever more important with the recent advancements in vehicle safety and efficiency. Modern vehicles' control systems have high technology embedded systems, which rely heavily on sensor data to control their stability and contribute to a safer driving experience. This contextual information from vehicles is fundamental to better understand traffic patterns, driver's behaviour and mobility patterns in a city. Car data can be gained via the Controller Area Network (CAN) bus, that can be queried through the automobile Sensing and Communication Protocols, On-Board Diagnostics (OBD).

2.1.1 Engine/Electronic Control Unit (ECU)

A vehicle collects information from hundreds of sensors that are connected to the Electronic Contro Unit (ECU) through an internally wired sensor network [37]. ECUs monitor and control many primary functions of the vehicles. Typically they are microcontrollers on new vehicles. They can be standard from the manufacturer, reprogrammable, or have the capability of being daisy-chained for multiple features. Tuning features on ECUs, allows the user to make the engine to function at various performance and economy levels. Some of the more common ECU's categories include:

- Transmission Control Module (TCM): handles the transmission including items like transmission fluid temperature, throttle position, and wheel speed.
- Engine Control Module (ECM): controls the actuators of the engine, affecting things like ignition timing, air to fuel ratios and idle speeds.
- Body Control Module (BCM): controls vehicle body features such as power windows and power seats.
- Vehicle Control Module (VCM): controls the engine and vehicle performance.

- Powertrain Control Module (PCM): controls the powertrain (a combination of an ECM and a TCM).
- Electronic Brake Control Module (EBCM): controls and reads data from the anti-lock braking system (ABS).

2.1.2 On-Board Diagnostics (OBD)

Nowadays, OBD systems are in most cars and light trucks. It is the language of the ECU. During the 1970s and early 1980s, manufacturers have started using electronic means to control engine functions and diagnose engine problems. This initiative targets primarily to meet Environmental Protection Agency (EPA) emission standards for reducing vehicle emitted pollution levels and for fighting engine failures for the useful life of the vehicle.

Through the years, OBD systems have become more sophisticated. OBD-II (described in 2.1.4), a newly introduced standard in the mid-90s, provides almost complete engine control and also monitors parts of the chassis, body and accessory devices, as well as the diagnostic control network of the vehicle. It is an expanded set of standards and practices developed by Society of Automotive Engineers (SAE) and adopted by the EPA and California Air Resources Board (CARB) for implementation by January 1, 1996. Hence, any vehicle manufacture since 1996 in Europe and the United States, is required by law to have the OBD-II computer system. Numerous studies rely on OBD data to determine driving profiles (e.g., [38][39][40]), FC (e.g., [41][42][43]) and measure the gas emission of a car (e.g., **[**44]).

2.1.3 Parameter Identification (PID)

The collected data from the sensors in the car, are available through OBD Parameter IDs or Parameter Identifications (PIDs). Typically, as a diagnostic tool, an automotive technician uses PIDs with a scan tool connected to the vehicle's OBD-II connector (section 2.1.6) to request data from a vehicle. The scan tool sends to the vehicle's bus which is defined in section 2.1.7, the code of a PID input by the technician (VPW, PWM, ISO, KWP, CAN (only after 2008)). The value for that PID is reported after being recognized by the specific device on the responsible bus for. Then the response is shown to the technician through the scan tool.

There are hundreds of sensors that can be accessed using PIDs. Not all PIDs are supported on all protocols and some of which are defined by the OBD standards and others defined by the manufacturers (not generally published) [45].

2.1.4 On-Board Diagnostics II (OBD-II)

An OBD-II interface is legally required in the United States since 1996, the year of manufacture. In Europe, it is legally required for cars with a gasoline engine built from 2001, for diesel vehicles built from 2003 and for trucks built from 2005. The system is usually employed to monitor the performance of some of the engine's major components, including those responsible for controlling emissions. It provides methods to query RT sensor values from the controllers and sensors of a vehicle, such as the ones related to the engine, actuators' state (steer angle, brake pressure, throttle, etc.) and exhaust system. Also, OBD-II interface (described in 2.1.6) is used to extract sensor data from the car's internal systems, which can be used to analyze and enhance concepts in the field of mobility such vehicle motion (speed, yaw rate, etc.) and environment.

When there's a problem with a vehicle, OBD-II protocol turns on the Malfunction Indicator Light or Lamp (MIL) (aka the Check Engine Light) on the dash (Fig. 3). Aftermarket maintenance services in the vehicles' repair shops, rely on the OBD-II scan tool to diagnose a vehicle for tracing the origin of mechanical problems [46]. A mechanic reads the Diagnostic Trouble Codes (DTCs) that is described in 2.1.5, of a vehicle using OBD-II interface to query those engine fault codes stored by OBD.

Fig. 3. Malfunction Indicator Light (MIL) on the dash [47].

2.1.5 Diagnostic Trouble Codes (DTCs)

DTCs are codes defined by the SAE, that the car's OBD system triggers in case an issue occurs in a vehicle by detecting a component or system, that's not operating within acceptable limits. Each trouble code corresponds to a detected fault in the vehicle, and hence it provides the technicians with information about malfunctions and their sources. These codes can either be generic or unique to the vehicle manufacturer. A DTC is made up of 5 digits, one letter and four digits such as P1234. Fig. 4 demonstrates the composition of a DTC.

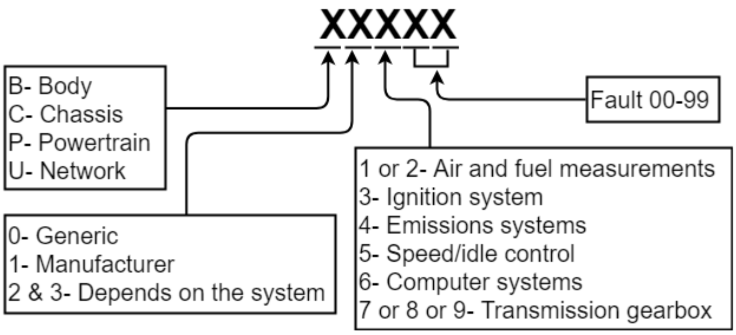


Fig. 4. OBD-II DTCs specification.

The following explains briefly the DTCs interpretations [48]. The **first letter** of the code marks the system (or the category) related to the trouble code. DTCs are categorized into four different systems;

1. Body (**B**-codes) covers functions that are, generally inside of the passenger compartment, providing the driver with assistance, comfort, convenience, and safety.
2. Chassis (**C**-codes) covers functions that are, generally, outside of the passenger compartment, including mechanical systems such as brakes, steering and suspension.
3. Powertrain (**P**-codes) covers functions that include engine, transmission and associated drivetrain accessories.
4. Network & Vehicle Integration (**U**-codes) covers functions that are shared among computers and systems on the vehicle.

The **first digit** in the code denotes if the code is a generic or manufacturer specific code.

- **0** for generic or global codes. It is common across most manufacturers, adopted by all cars that follow the OBD-II standard.
- **1** for manufacturer-specific or enhanced fault codes. It means that they are unique to a specific car make or model, not to be used generally by a majority of the manufacturers.
- **2** or **3** in case the type depends on the system. B2xxx and C2xxx codes are controlled by manufacturers, while B3xxx and C3xxx codes are reserved. P2xxx codes are generic codes while P3xxx codes are manufacturer controlled. U2xxx codes are manufacturer controller as well as U3xxx codes.

The **third digit** indicates what system the trouble code references:

- **1** or **2** for air and fuel measurements.
- **3** for the ignition system.
- **4** for the emissions systems.
- **5** for the speed/idle control.
- **6** for the computer systems.
- **7** or **8** or **9** for the transmission gearbox.

Digits **four** and **five** show the specific failure code, ranging from 00 to 99. These are based on the systems that are defined with the third digit.

2.1.6 OBD-II Output Measurement Mean – Data Link Connector (DLC)

In-car sensor data can be accessed externally using a universal interface, the OBD interface (or OBD port, also known as Data Link Connector (DLC)) for reporting and diagnosis purposes [35]. The most recent interface is the OBD-II, which was introduced to standardize the physical connector, its pinout, the signalling protocols and the format of the messages they deal with.

Fig. 5. OBD-II adapter plugged into car's female 16-pin OBD-II connector [50].

The OBD-II interface is located in the footwell - in most vehicles directly underneath the steering wheel under the dash, near the driver's seat, or in the vicinity of the ashtray. It is easily accessible from the driver's seat without the use of tools to access it (Fig. 5). They should be a note on a sticker or nameplate under the hood: "OBD II compliant".

To request information from ECU, a software application (installed on a desktop or an Android platform) should be connected via Bluetooth to the vehicle OBD-II interface through an OBD-II adapter scanning tools. This OBD adaptor is the link between the diagnostic connector and the device that runs software for reading codes and data.

2.1.7 OBD-II Protocols

There are five basic OBD-II signalling protocols in use, allowed with the OBD-II interface (Table 1). Each of those protocols is with minor variations on the communication pattern between the OBD computer and the scanner console or tool. All of them use the same OBD sixteen pin connector (eight pins at the top row and eight pins at the bottom row) as depicted in Fig. 6.

Table 1. ODB signalling protocols.

Protocol	Transfer Rates
SAE J1850 Pulse Width Modulation (PWM)	41.6 kbit/s
SAE J1850 Variable Pulse Width (VPW)	10.4 kbit/s
ISO 9141-2	10.4 kbit/s
ISO 14230 Keyword Protocol 2000 (KWP2000)	10.4 kbit/s
ISO 15765 Controller Area Networks (CAN)	250 or 500 kbit/s

Pin 16 is used to supply power via the car's battery - often also while the ignition is off. Pin number 1, 3, 8, 9, 11, 12 and 13 are blank for vendor options. Depending on which pins are populated for the remaining pins in the DLC, the used protocol in the vehicle can be known (Fig. 7) [51]. For instance, for a protocol to be:

- J1850 PWM, pin 2 and pin 10 must be there in the connector (the connector must have metallic contacts inside pins 2, 4, 5, 10 and 16).
- J1850 VPW must have pin 2 (the connector must have material contacts inside pins 2, 4, 5, and 16, but not 10).
- ISO9141 and 14230 (KWP2000) must have pin 7 while pin 15 is optional (the connector must have metallic contacts inside pins 4, 5, 7, 15, plus 16).
- ISO 15765 (CAN) should have both the pins, 6 and 14 (the connector must have material contacts inside pins 4, 5, 6, 14 and 16).

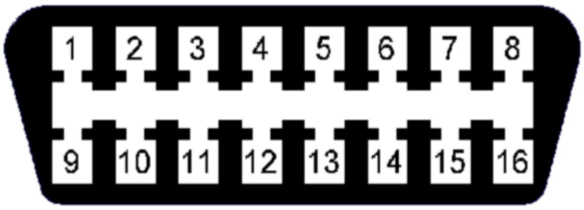


Fig. 6. Female OBD-II 16 pin DLC and pinout [52].

As a rule of thumb, Chrysler products and all European and most Asian imports use ISO 9141 circuitry or KWP2000. GM cars and light trucks use SAE J1850 VPW (Variable Pulse Width Modulation), and Fords use SAE J1850 PWM (Pulse Width Modulation) communication patterns. Controller Area Network (CAN) described in 2.1.8, is the newest protocol added to the OBD-II specification, and it is mandated for all 2008 and newer model years [49].

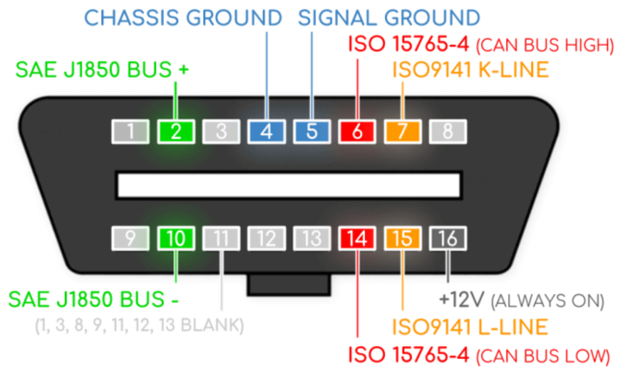


Fig. 7. OBD-II DLC male to female cable pinout instruction [51].

2.1.8 OBD-II Controller Area Network (CAN) Bus or (ISO 15765)

CAN is one of the transport protocols of the OBD-II specification. The CAN Bus system was developed by Bosch in 1986. It is a vehicle bus standard designed to allow ECUs (e.g. brake, engine, electronic fuel injection, automatic gearbox, ABS) to communicate with each other within a vehicle without central computer [53]. CAN bus is designed for sending many signals over a few wires for greatly reducing the vast amounts of wiring. It is a dual-wire network that allows the ECUs to send data back and forth (such as vehicle speed, voltages, engine coolant temp, throttle position and switch states).

A. CAN Bus System Schematic

On the physical layer, CAN bus consists of two dedicated wires for communication; CAN high and CAN low. The CAN controller is connected to all the components on the network via these two wires. Both wires are linked to the vehicle's DLC for diagnostic purpose. ECUs communicate via a single CAN interface. This makes the control system to be easy to operate or regulate and reduces the errors, vehicle weight and costs (Fig. 8).

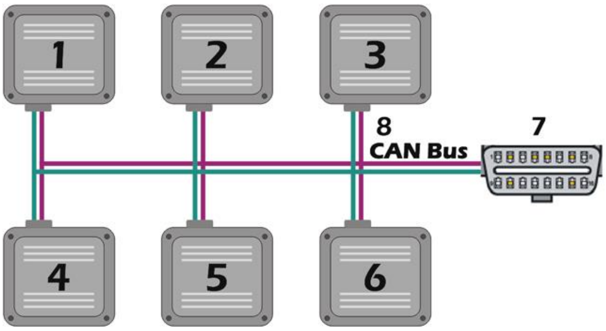


Fig. 8. CAN bus network. ECUs CAN nodes (1-6). OBD DLC (7). Two CAN wires (8) [54].

B. Principle of operation of the CAN BUS

Each network node has a unique identifier. When an ECU sends a message, all other ECUs on the bus receive it in parallel. A node only responds when it detects its own identifier (Fig. 9). This allows for multiple modules to share inputs and data without the need to run multiple wires to each.

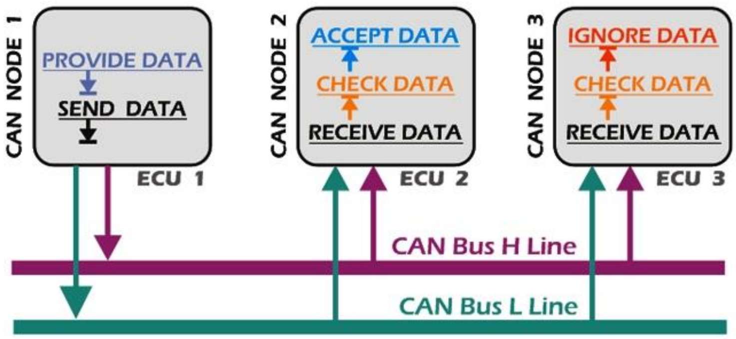


Fig. 9. CAN bus communication [54].

These signals are encoded into messages which store all driving events on an onboard recorder from where data is retrieved and stored in a database file (.dbc extension), which stands for database CAN. If two messages are being sent at the same time, the message with the higher priority is sent first and the other messages back off and wait. For instance, the brake may have the highest priority. When applying a brake, the car has to slow down as soon as it can.

When the CAN bus is in idle mode, both lines carry 2.5V. On the other hand, when data bits are being transmitted, the CAN high line goes to 3.75V and the CAN low drops to 1.25V. Thereby, generating a 2.5V differential between the two lines (Fig. 10). Since communication relies on a voltage differential between the two bus lines, the CAN bus is not sensitive to inductive spikes, electrical fields or other noise. This makes CAN bus a reliable choice in an electrically noisy environment such as a vehicle.

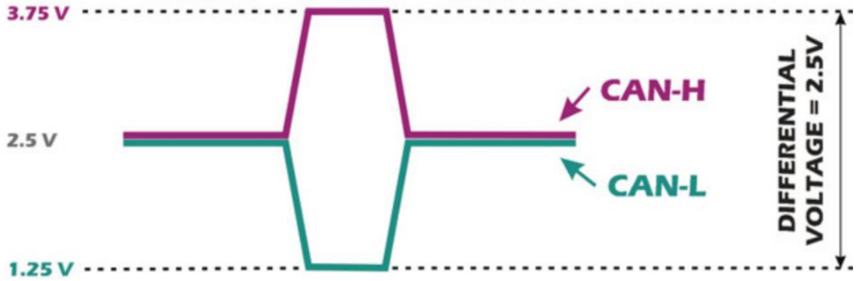


Fig. 10. Stable Differential value (2.5V) between CAN bus two lines [54].

Numerous studies have plunged into driver style assessment and categorization based on FC. A promising and effective approach is to use eco-driving (economic or ecologic driving) or “eco-friendly” technologies, that monitor the driver behaviour [3][8]. The concept of eco-driving represents a driving culture with smarter and safer driving towards pollution reduction contribution. It encourages to fuel-efficient driving leading in carbon dioxide (CO_2) emissions reduction [16]. Furthermore, this concept offers benefits towards environmental protection apart from fuel-efficiency, such as cost-saving, greater safety, riding comfort, and less noise pollution [40][56].

This approach is defined as a decision-making process based upon on understanding of what primarily affects FC, including recommendations on person’s driving attitude towards decreasing acceleration, braking, idling, speeding and driving aggressively [57]). Studies have proved the effectiveness of eco-driving support system in saving fuel (e.g., [12]).

2.2 Eco-driving Techniques and Recommendations

The concept of eco-driving comprises a set of techniques designed to improve fuel-efficiency and reduce environmental damages that contribute to the effect of greenhouse gas (GHG) emissions resulted from the transport sector. It also includes recommendations on vehicle maintenance – the way and frequency a vehicle is used, its configuration and accessories (e.g., managing engine load and check tires pressure). Eco-driving techniques can be summarized as the following:

(1) Hypermiling: this approach compromises a collection of driving methods which aims to improve a car's fuel economy by reducing the demands placed on the engine [58].

   a. Coasting: It encourages to drive avoiding the use of brakes as possible – to coast as much as possible. It often occurs when travelling downhill (with the engine off and drafting) or while turning a corner or emerging from a junction, either in neutral or the clutch is down [59].
   b. Route selection and trip timing: Choosing a route (e.g., smooth road surfaces, take the road less travelled, avoid peak traffic) allows to maximize the fuel economy.
   c. Trip-chaining: It refers to visit multiple destinations sequentially or grouping the trips to benefits from a warm engine instead of making multiple separate trips [60].
   d. Carpooling or ride-sharing: This means that several people travel in a car together and share the cost of the ride. Typically, each person takes a turn to drive the others in an arrangement. For instance, 'BlaBlaCar' allows drivers who have empty seats to find carpool partners, by connecting people who need to travel [61].

(2) Cruise control is known as speed control or "autocruise": It is a system for controlling automatically the car' speed on highways [62].

2.2.1 CO_2 Representative of Emissions from Transportation

The consumption of fuel is highly correlated with CO_2 emissions [57]; those emissions are in a linear relationship with the fuel mass that is combusted in the engine [26]. This is verified with my data analysis in section 3.5.1 (Fig. 21).

However, FC is less correlated with other harmful vehicle emissions; carbon monoxide (CO), hydrocarbon (HC), and mono-nitrogen oxides (NOx) [63]. Adding to that, the selection of CO_2 as the representative of emissions is due to that it is more concerned by drivers and the general public [64].

2.2.2 Eco-driving and Driving Analysis Rules

There are five common eco-driving rules, which are guided in the developed countries. Those rules have been approved for having a significant effect on vehicle' FC and emissions [43][65], mainly;

(1) avoid rapid start and accelerate smoothly,

(2) decelerate smoothly by releasing the accelerator in time,

(3) maintain a steady speed by anticipating traffic flow,

(4) shut down the engine for long stops,

(5) shift up as soon as possible and avoid engine revolutions at a high level.

Typically, it is recommended to be gentle with the accelerator, with no more than half throttle [66]. Pressing the accelerator pedal pushers more fuel into the engine, making it run faster. This lowers the fuel economy level and hence increases the output of pollutants. [63] suggested avoiding extreme acceleration, acceleration with strong power demand, acceleration with moderate power demand, and engine speeds over 3500 RPM.

Furthermore, the current gear engaged is considered valuable information to describe a driver’s habits [67]. Current gear information is not available in any signalling protocol of the OBD interface in cars with manual transmission. Hence some studies work on current gear detection for manual cars to assess driver’ style upon analyzing vehicle’ historical sensor data. For example, [25] proposes a gear shift recommendation service of the best gear, based on a linear relationship between speed and RPM, after applying a clustering algorithm considering speed and torque. Their finding reaches up to 29% average of fuel-efficiency and up to 21% averaged in CO_2 emissions reduction. For more fuel economic, it is recommended to shift at lower RPM, between 2000-3000, staying less than 3500 RPM [57].

[63] has investigated urban driving pattern parameters that have the greatest effect on emissions and FC through regression analysis, such as speed, acceleration and gear-changing. The analysis depended on the environmental impact such as the route choice. It was found that speeding in itself does not influence emissions considerably, and the attempt to lower speed limits for traffic safety reasons are not key problems from an environmental point of view. Instead, the main focus should fall on adapting the environments and driving in a way that discourages heavy acceleration, power demand, and high engine revolutions.

2.2.3 Eco-driving Coaching Studies

Some companies have recognized since long the value of adopting the eco-drive approach. Thus, they are investing in specialized training services to teach their employers how to drive efficiently. They analyze their drivers’ behaviour before and after given such eco-driving sessions. Moreover, they rank their drivers upon the best practices. Such training has been reported to reduce FC between 2.6% [68] and 20% [69], in real-world driving conditions.

[56] investigates the impact of providing RT eco-driving advice to drivers based on RT traffic speed, density and flow with a reducing of FC in 10-20% without a significant increase in travel time. [65] characterized an efficient driving process for companies of the road transport sector, ranking each driver

after an individualized learning process to reduce FC with a low investment. CGI Group Inc [70] conducted a study to compare the impact of eco-driving coaching across seven European Union countries based on more than 3 million Scania truck trips. Also, they proposed an estimated effect of coaching (EEOC), which provides a realistic estimate of the fuel savings to be gained from eco-driving coaching. A conducted study in Australia showed a reduction of 4.6% in FC with drivers who received eco-driving instructions after monitoring 1056 private drivers (853 drivers 0received education in eco-driving techniques) over seven months [71]. A similar study was conducted in Canada, showing a decrease in FC and CO_2 emissions up to 8% [72].

2.2.4 Eco-driving Training Types

Eco-driving training mainly includes (1) static and (2) dynamic types.

(1) The static approach aims at urging drivers to apply general eco-driving techniques after learning through brochures, websites and other information sources.

(2) The dynamic eco-driving instructions involve providing drivers with direct feedback [73], which may be carried through three sensory modalities: visual, auditory, and haptic [74].

Also, the strengths and weaknesses of each type of eco-driving feedback were compared in previous studies (e.g., in [75]). For example, eco-driving onboard devices (such as vehicle dashboard or smartphone applications) and offline feedback were usually used to influence driver behaviour through visual and auditory feedback. They are easier to apply then haptic means, but they might affect negatively driver' attention and performance [76][77].

Eco-driving feedback includes conventional dashboards, hybrid vehicle dashboards, smartphone applications, offline feedback systems, dedicated aftermarket feedback devices and haptic pedal feedback [75]. By using these eco-driving systems and devices, drivers can know their relative levels of driving with respect to FC and emissions, and also gain tips to improve driving skills [78]. The experiments demonstrated that providing feedback was effective in saving fuel and reducing emissions to varying degrees [79][80][81].

As an example, [12] has developed an eco-driving feedback system to support eco-driving training upon the five common rules of eco-driving in developed countries (stated in 2.2.2) since they are closely related to driver performance while driving and these behaviours are also easy to measure in the simulator. The simulator provides a second-by-second calculated vehicle' FC by the carbon balance method [82] and emissions model based on vehicle-specific power distribution [83]. The results showed a reduction of 5.37% for CO2 emissions and 5.45% for FC. This was achieved by providing both (1) dynamic RT voice prompts once non-eco-driving behaviour appeared, with RT visualization for CO2 emissions curve, and (2) static feedback with the mean of eco-driving evaluation report after driving (including drivers FC rank, the potential of fuel-saving and driving advice corresponding to drivers driving pattern). The five observed behaviour conditions are;

a) Accelerate: equal to or greater than 3 m/s^2 and this state lasts 3 or more seconds;
b) Decelerate: less than or equal to -3 m/s^2 and this state lasts 3 or more seconds;
c) Shift gear: (gear = 1 and speed >30 km/h), or (gear = 2 and speed >40 km/h), or (gear = 3 and speed >50 km/h), or (gear = 4 and speed >60 km/h);
d) Vehicle Speed choice on freeways: the vehicle speed is more than 110 km/h;
e) No-idling: the vehicle is stopping but the engine is running and this state lasts one or more minutes.

2.2.5 Controversial Issues in Eco-driving Definition

The definition of eco-driving varies widely from source to another [57]. Sometimes, the recommendations are restricted to a driver's operation of the vehicle [6], but occasionally it includes vehicle purchase and maintenance decisions (e.g., [3]).

Some researchers claim that eco-driving characteristics are misleading with a general definition [6]. The research in [57] has been targeted to make significant progress toward a comprehensive and precise behavioural typology of eco-driving by synthesizing, clarifying, and re-framing prior definitions. It has been shown that there are imprecise, inconsistent, and incomplete behavioural topographies and also behavioural functions are conflict in prior relevant studies. This makes difficult to decide the potential of the overall eco-driving savings and then to refine the corresponding effectiveness of various interventions on different eco-driving comportment. The typology of [57] is based on the concept of behavioural function, proposing the following six eco-driving behavioural classes:

1. Driving (e.g., operating the vehicle to control direction and speed)
    a. Acceleration:
        - Pedal: It is recommended to be gentle with the accelerator, with no more than half throttle [66], since it pushes more fuel into the engine, making it run faster and also lowering fuel economy.
        - Gear: Shifting up to higher gears early is widely recommended.
    b. Cruising: 'Obey speed limits' is preferable to 'don't speed' as sticking to speed lesser than the speed limit is not just obeying the traffic law but also ensuring safety; Minimize load on the engine (referred as air and fuel in the cylinders). As the gas pedal is pressed, the throttle opens slightly more, increases engine speed. Increasing speed requires acceleration of the vehicle, which increases the load of the engine. For more fuel economy, a steady speed (of 100 km/h on

freeways) has to be maintained; e.g., the use of cruising speed on motorways delivers a positive effect on FC and thus to CO_2 emissions.

c. Decelerating: Minimize acceleration and braking (make progress without using power or as little effort as possible), and maximize coasting (2.2).
d. Waiting: Reduce idling (the vehicle is stopping but the engine is running) by keeping out of congested areas and shutting down the engine when the stopping time is, or expected to be, more than one minute.
e. Parking; In cold weather, park in a warm place (e.g., garage) so the engine does not begin as cold. While in warm weather, use a sunshade or parking in the shade for keeping the vehicle cooler. This reduces the need for air conditioning and also minimizes evaporation of gasoline.

2. Cabin comfort: Minimise extra energy loss that costs fuel and money.
   a. Comfort: Optimize the use of air conditioning and rear-window defroster, close windows at high speeds; Slow down the engine warming process by substituting the use of cabin heater with other means such as electrically heated seats, electric front window defroster, or electric heat pump system.
   b. Communications: Switch off electrical entertainment equipment if not needed. Moreover, turn off air conditioning, heating, and auxiliary electronics before shutting off the engine to decrease the engine load when starting the next trip.
3. Trip planning: Select travel time and routes (road type, grade, right turns, congestion, trip-chaining).
4. Load management: Manage cargo weight to maximize aerodynamics. Avoid unnecessary weight and remove roof racks. Necessary cargo should be stored inside the vehicle whenever possible, since hauling cargo in roof racks and boxes adds wind resistance, which leads to an increase in FC.
5. Fuelling: selecting proper grade/octane of gasoline or renewable fuel. It also involves fuel evaporation, that can be minimized by refraining from topping off and make sure the gas cap is intact. It is recommended to fuel at night and parking in the shade to prevent evaporation.
6. Maintenance: Change engine oil every 3000–5000 miles as clean oil results in better fuel economy, select oil, check tire pressure regularly and keep tires properly inflated (tire pressure deterioration status influences directly the level of FC), select tires.

- **Economical versus Ecological Driving:**

In the literature, there is a conflict between the most economical and the most ecologic driving style. For example, driving at the lowest possible cruising speed in the highest possible gear is the most fuel-efficient, but this requires high torque engine operations that result in greater HC and CO emissions [57].

- **The Tradeoff between Eco-driving and Safety:**

The proponents of eco-driving often recommend accelerating ‘moderately’ without further operationalization (e.g., [84]). This encourages for ‘safe driving’ with avoiding sudden accelerations, rapid braking, and cornering actions [85]. Conversely, ‘aggressive driving’ or driving not safely, is sometimes used as a catch-all for vehicle operations that are not eco-driving, such as hard acceleration and braking, excessive speed, open windows, etc. [86].

However, considering safety in the eco-driving definition is a controversial issue, since the fact that unsafe fuel-saving comportments will be reinforced any time fuel-saving in general [57]. For instance, there is a contradiction between the eco-driving and hypermiling functions, which includes unsafe tactics, such as coasting while driving. With this approach, the vehicle goes into a ‘free-wheeling’ state, letting the vehicle to move not using the engine, hence not under the driver’s control. Going along downhills with either in neutral or the clutch is down or the engine off and drafting, is potentially dangerous. Because this leads to less driver’s control over the vehicle since the vehicle is in a ‘running away’ state – gaining speed quickly. This can be frustrating and difficult if other drivers keep cutting in front of the driver, requiring him/her to brake far harder than normal. Furthermore, coasting while turning a corner or emerging from a junction situation, usually forces the driver to turn too wide. This situation puts the driver and surroundings in danger with ongoing traffic [58][59][87].

- **Noise Emissions Challenge:**

Eco-driving definitions encourage reducing noise emissions. However, the estimation of noise emission is not straightforward since many parameters such as road and tyre quality, or aerodynamic drag influence noise emissions from vehicles [88]. Nonetheless, the enviroCar project [26] uses OBD-II sensors to support the estimation of local noise emissions by measuring the engine speed. Higher engine speed correlates to higher noise emissions emitted by the tyres [89].

Furthermore, on some occasions, drivers adopt a less efficient driving style than the one they previously had, if they misinterpret the eco-driving advice [10]. Hence, there is a need for easy understandable and incentive practical advice.

2.3 Gamification

Since 2008 different attempts have been made to define 'gamification'. In this section, I describe shortly this term and its impact on engagement and educational scenarios with SGs. Gamification term refers to the application of game-style mechanics and experience designs in non-game contexts and activities, to digitally engage and encourage (individual or group) participation with positive behaviour motivation to achieve a more preferred outcome [90][91]. This definition involves both types of gamification; (1) structural gamification ('game mechanics') and (2) content gamification ('experience design') and also highlights the importance of engagement and users' intrinsic motivation.

Gamification injects fun elements into applications by making potentially tedious tasks more fun to incentivize the users toward goals' achievement. Gamification's motivation toward further players' improvement is attained via virtual gaming reward mechanics (e,g., points, levels, awards, leaderboards or badges) to visualize the progress towards a certain goal. Inserting those gameplay elements by designers within a gamification platform into existing non-gaming settings, leads to high levels of player's engagement and a willingness to return to a product or service on numerous occasions.

According to [92], a gameful design can motivate positive involvement. Typically, users enjoy challenges for being rewarded. Gamification is effective to increase engagement since it leverages people's natural desires for status, achievement, competition and to be part of an inclusive social community. Many fields rely on gamification to increase participation and improve productivity e.g., among the academic world, where teachers exploit the use of gaming scenarios towards adopting students' skills while playing [93].

Yet, gamification is often an essential feature in apps and websites designed to motivate users to meet personal challenges, like weight-loss goals and learning foreign languages. Also using gamified elements in business, users' attention and loyalty can be recognized while creating an identity for their brand or product. Managers believe that gamification is useful for both customers and employees together to motivate their engagement and improve some of their behaviours [93]. Businesses such as IBM have already recognized gamification as an enabler, beneficial and practical tools for training. IBM "Business Process Management" (IBM) has developed "Innov8" game to educate and train new employees, addressing numerous topics with the virtual environment [94]. It is used academically by the University of Southern California. Gamification motivation was used toward ecological attitude in [95]. Users earn points when performing ecological actions which are shown by their friends on the website's timeline. Also, users' activities may be shared with other external social networks.

Yet automakers of some vehicle models that target efficient driving, are using gameful design [96] – with a strategy based on the use of colour and contrast that helps reduce time glance – to supply the drivers with virtual rewards presented with simple gaming interfaces (e.g., trees, flower or medals) based on their eco-driving achievements. For example for the hybrid models in Ford's 'SmartGuage' with 'Ecoguide', there is a functionality to provide RT feedback about driver' habits that includes fuel and battery power levels and average miles-per-gallon with a rich-colour LCD screen (with non-distracting animation) [97]. It was launched to the U.S. market in March 2009 for the model year 2010. Another example is with Honda's Eco Assist which rewards 'green' driving styles by the number of leaves displayed in the instrument cluster, indicating the level of driving performance from the environmental side. The system debuted on the Honda Insight hybrid in 2009 and it has expanded the availability to other Honda models including the Civic, CR-V, and Accord [33]. To the best of my knowledge, the technological details of the manufacturer-specific solutions towards efficient drivers are not disclosed in the literature.

Following this industrial direction, I intend to support the development of new gameful solutions for fuel efficiency improvement, through a tool available to all the drivers, by using the standard OBD-II vehicular interface and exploiting ML and AI algorithms for extracting RT information about the drive and to support coaching.

There are two basic ways to maximum mpg – have a fuel-efficient vehicle and drive the vehicle efficiently. For the new vehicle models, those two points are considered by performing two basic functions. For instance in Honda's Eco Assist, there is a sophisticated feedback system for coaching drivers in fuel-efficiency (e.g., via colour-changing displays for indicating fuel-efficiency in RT, typically integrated into the speedometer,) that coaches drivers to develop a driving style towards more fuel-efficiency driving. Hondas are known for their great fuel economy, but also, there's an 'ECON' button. By pressing this button, the vehicle switches into a super-efficient mode that automatically makes slight adjustments to maximize fuel-efficiency. This is achieved by lightening the load on the engine so the consumed fuel is mostly being used for driving power. Some of its actions are, a more gentle throttle is maintained; shifts into higher gears more quickly than without the 'ECON' mode for vehicles with an automatic transmission or indicate when to shift gears for vehicles with a manual transmission; limits the running time on the air conditioner or heater [33].

2.3.1 Serious Games (SGs)

A SG or applied game presents a step beyond gamification. Whereas gamification takes elements, e.g., point scoring, to improve aspects of the experience, e.g., motivation, it's not a game in itself. In contrast, SG is designed to be a game that is not solely entertainment-based. SGs present a promising approach to training and learning content delivered in a game-based environment more than entertainment, aiming at user motivation and behaviour improvement [98]. Thus in such games, users can learn new habits while having fun.

These proactive coaching and training methods in a digital gaming environment – learning while playing – is known as game-based learning. Their use has grown in several sectors such as education, defense, aeronautics, science and health. SGs include educational games or edugames, simulations, healthgames, newsgames, advergames, e.g. “games built to promote products or services” and games for change, e.g. SGs for some social benefit or to improve driving style. There are many purposes for such games, e.g., teaching mathematics, practicing a language, training firefighter crews in emergencies and training a sales team.

As one of the SGs examples is ‘Dragon Box Elements’ designed for kids of nine years-of-age and upwards [99]. It is one of the simplest (and effective) games in the field of education, where the students learn Math while they have fun with a video game. Players have to build an army, defeat the evil dragon ‘Osgard’ and save Euclid’s island. This requires them to learn basic geometry and the theorems of Euclid himself to proceed.

Many SGs have been developed to be used in class, or a safe area. Realistic 3D simulations allow high fidelity training in environments and situations in which real-world training would be costly, dangerous or unfeasible. Moreover, synergy with the IoI technologies [100] has a promising potential, especially for SGs related to field operations. The diffusion of low-cost IoT devices is enabling a new generation of games related to field operations that are implemented to work in the wild [101]. For instance, in the road-traffic area, there are simulation-based games supporting learning and practicing [102], while a new generation of IoT- enabled games has arisen supporting RT interactions in the application situation [103] (e.g., driving game on smartphones equipped with inertial sensors [104]).

High-quality SGs are typically designed to transfer knowledge and skills, from game-play to real life [21][105]. This might be achieved by inserting game elements in real-world processes (e.g., through gamification [106]), or in “reality-enhanced” gaming approach – a user’s real-world activities do feed a digital game [21][107][22]. Such “reality-enhanced” games (or, more specifically, SGs) are a specialization of pervasive games [108][109][24], a game genre in which players are immersed in real-life situations and leverage new types of contextual interactions therein (allowing user interaction with a virtual environment) [110].

2.3.2 Pervasive Games

In pervasive games, the game goes beyond the bounds of one screen bridging the physical and the digital worlds. In such games, the players are immersed in real-life situations. This is achieved by leveraging new types of contextual interactions therein by adding interesting dimensionality to increase the immersing of the players [110]. This is usually by combining elements from the real-world and computational/virtual modules taking place across multiple devices or range through (pervade) the real world [108]. Recent developments in pervasive gaming include integration with virtual reality (VR), augmented reality (AR) and mixed reality (MR). Pervasive games are very relevant also for the design of SGs [111].

2.3.3 Reality-Enhanced Serious Gaming (RESG)

Reality-Enhanced Serious Gaming (RESG) term refers to a specialization of pervasive adaptive training gaming, where games are fed with data collected from the field in RT, in the order of a few seconds, with negligible computation latency assessment to provide input to the game toward player experience motivation with the actual activities [21].

In such games, in-game progress is due not only to the digital gaming ability of a player but also depends on sensing a user's performance in the actual target field, which is then transformed into a proper game mechanic (e.g., score, energy, etc.), according to the actual SG logic. This is an evolution of pervasive gaming [24], where the game's fictive world blends with the physical world connecting a digital game environment with reality, and allows opening and exploiting a direct, possibly RT, link between a game and a training objective. Hence, field users' performance becomes a key factor and should be easily understandable to supply RT effective coaching feedback to players for such kind of games, which are typically also played in the field or the wild [22].

2.3.4 Driving Behaviour and Gamification

Gamification approach has emerged in the automotive industry for its motivating and inspiring potential [19][20]. Studies (including mobile enterprise) have used gamification tools to score behaviour of drivers (e.g., 1 to 100) and "gamify" it to rank performance against their peers. They find the errors of drivers during each drive using multiple criteria.

Some of those games, provide RT coaching and training based upon data extracted from in-vehicle or smartphone or combining both sources. They also trigger warning in case aggressive driving behaviour is detectable such as speeding, harsh acceleration, hard braking and cornering. Those instant tips help drivers to enhance their driving styles while driving, so that learning opportunities are immediate and not reactionary. For instance, for the administrators in fleet management domain who are keen on reviewing their drivers' driving patterns, those games can help them to minimize fuel use and excessive idling by monitoring when and how an asset is being used through RT driving data of their drivers.

Yet, studies have relied on gamification's motivation that involves incorporating elements such as scores, leaderboards, and badges to prompt drivers towards more fuel-efficient and safer driver behaviour. Typically, human has innate desire to win and to beat other peers' score. Those games' motivation elements encourage continuous improvement, incentivize the driver to be recognized as a leader shapes operator driving behaviour over time, which then leads to safer and improved performance.

In-car gaming is gaining relevance, with opportunities – it has the potential of making use of all the cool properties of the car itself, the practices of driving, and of shared driving experience and challenges (e.g., aspects of driving, such as driving as an arena for gaming) [112]. Whilst driving and travel-related SGs can have a range of objectives such as promoting more low-carbon vehicle use [100] and encouraging the use of different transport mode and route choices [113]. Several in-car game concept designs have been discussed in [114], also considering driving style. In [115], an incentive-based mechanism was adopted to improve driver behaviour in managing traffic congestions. [116] evaluated the effects of gamification on driving, especially considering boredom. The Car-wings application by Nissan, represents FC' status vs money spent and also provides a comparison among different driver performance [117]. The Car2Go application provides gamification features to support environmental friendly behaviour [118]. [119] discussed the effects of gamifying recreational bicycle riding, with positive and negative consequences.

[103] has presented a Windows Phone application to report road accidents using a game layer to motivate drivers (e.g., through points, levels, challenges). Some motivations for fuel-saving were achieved by combining gamification with social networks. [120] has presented an approach for utilizing FC' data in an incentive system for the Tampere City Transport company, based on comparing individual driver's average of FC with the average FC of all drivers in a specific group (formed with similar vehicles, routes, and time of day). [121] has presented a social awareness system to promote eco-driving and safe-driving by implementing some social experiments on a website through communication technology for gathering information about the driving styles using both GPS and motions sensors. [10] implemented an awareness game to encourage drivers to save fuel using some eco-driving tips. It compares the vehicle telemetry with other users with similar characteristics. Scores were assigned to users from the energy efficiency point of view, and then users are grouped for comparison based upon: time, average speed, the percentage of time that the vehicle was stopped, the percentage of driving time at more than 50 km/h, the stop rate, the traffic events, and the weather conditions. Drivers can share their scores with other users or on social networks. The results of their experiments, recorded on three different routes by 36 drivers, show that gamification tools and eco-driving assistants help drivers to not lose interest in fuel saving.

2.3.5 Gamification Distractions and Safety

Given the frequently critical operation context, user experience should be friendly, usable, beneficial and attractive towards the game's sustainability for retaining and gaining users. Also, serious implications, in terms of distraction and citizen privacy violation [122] should be carefully taken into account in the design and deployment of such games. By no means, games should become an instrument through which citizens are persuasively and "nicely" controlled. Drivers need to not be visually distracting with the game design or engagement with long glances towards the smartphone/vehicle screen. They should keep paying more attention to the road ahead focusing on their surroundings (pedestrians and other vehicles) rather than on the phone or the screen of the car where the driving game is integrated. For instance, a game' design who advises drivers for speeding control regardless of road distraction. The experiments in [116] suggest that "it would be better to drive 1 km/h too fast than to be distracted from the road".

Gamification distraction may be reduced by decreasing the time of focusing on the game screen to follow the driving evaluation based on the driving progress. A prudent balance between the driving tasks and the visual distraction (via game advice engagement for improving the driving tasks) need to be taken into account. This can be bypass by studying the position of the smartphone that always impact visual distraction [116]. Also, providing driving advice via ambient voice prompts (or haptic feedback) instead of visual text is another option [123][124]. Furthermore, a gameful design needs to be carefully designed and evaluated. It should be self-descriptive as possible for reducing time glance. For instance, some driving companies (e.g., the hybrid models in Ford's 'SmartGuage' with 'Ecoguide') have followed a strategy based on the use of colour and contrast to supply the drivers with virtual rewards presented with simple gaming interfaces (e.g., trees, flower or medals) based on their eco-driving achievements (section 2.3). Moreover, while driving, the recommendations should be easily understandable and incentivise practical actions avoiding driving distractions and irritations (check the part "The Tradeoff between Eco-driving and Safety" in 2.2.5).

2.4 Anticipated Review of the Used Techniques in the Methodology Section

This section introduces the four techniques that I have used in the methodology (described later in section 4); fuzzy logic (FL), random forest (RF), Support Vector Machine for Regression (SVR) and Artificial Neural Networks (ANNs) which are introduced in sections 2.4.1, 2.4.2, 2.4.3 and 2.4.4 respectively.

2.4.1 Fuzzy Logic (FL)

FL is capable of transferring human knowledge and expertise into a mathematical model through if-then rules by matching any set of input-output data. It is used to monitor non-linear systems that are difficult to deal mathematically. Unlike classical control strategy, which is a point-to-point control, FL control is a range-to-point or range-to-range control. There is no absolute truth about the state of the studied characteristic. The concept of linguistic variables and FL has been reported first time by "Lotfi Zedeh" [125]. It provides flexibility and simplicity to complex real-world problems containing uncertainty, imprecision, non-linearity and granularity (incomplete information) [126].

Unlike some other data analysis techniques (e.g., Neural Networks), a FL model itself is readily understandable, which is a key requirement to give coaching feedback to drivers. It uses natural language techniques and variables which are based on the degree of truth, that are easier to understand for human beings and inference process [126][127].

FL has been frequently applied to complex real-world problems with incomplete, imprecise and non-linear data, combining flexibility and simplicity [28][126]. It has been considered "promising" in a recent review of driving style analysis systems [85].

FL has been applied in numerous application domains, including driver behaviour analysis. [128] discussed a vehicle speed limit model with a road safety model based on fuzzy rules. A recent review paper has shown the relevance of FL for driving style analysis [85]. In addition, this technique has been used for driver fatigue and distraction identification (e.g. in [129]), scoring (e.g., in [130][131]), driving style recognition (e.g., in [39][132][133][134][135]) and FC estimation (e.g., in [28][43]). TPS, RPM and car speed – that are my selected sensors - have been used in [43] to develop an estimation system, that outputs a categorical level of FC (very low, low, medium, high and very high).

2.4.2 Random Forest (RF)

Among ML techniques, an ensemble learning approach – the process of generating a team of ML models and then aggregating their results to obtain much powerful predictive performance, has gained a great interest nowadays. One of the most effective ensembles learning approaches is the RF supervised technique, carrying prediction and classification problems, and it's also applicable for time series analysis.

Fig. 11 depicts the general architecture of a RF model. In RF, several (decision or regression) trees that construct a forest, are built and trained to empower predictive models with global more accurate and stable predictions. The prediction is obtained as a majority vote over all the trees in the forest for classification or as the average of outputs by different trees in the forest for regression [136][137]. Each tree is trained on a bootstrap sample, and optimal variables at each split of the tree's nodes are selected from a random subset of all available variables, which increases the purity of the node and results in the most homogeneous sub-nodes [138].

This strategy enables RF to be robust against overfitting and be held as an outstanding predictive model amongst many classification and regression tasks [139], including discriminant analysis, support vector machine for regression (SVR), and neural networks (NNs) [140]. RF methods are reasonably fast in learning [140] and can be easily parallelised if more speed is required. In addition to prediction, RF is also considered as a dimensionality reduction and feature selection method, where it helps in the interpretation of important informative predictors for the outcomes.

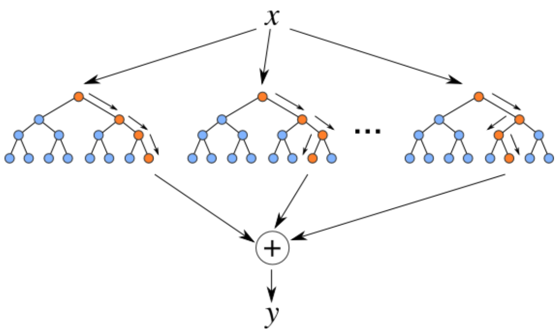


Fig. 11. Example of general random forest architecture [141].

RF has been used in a myriad of domains. For instance, [142] used RF to forecast droughts, and the study demonstrated that RF outperforms Autoregressive Integrated Moving Average. Also, RF has been applied in FC prediction. For example, it was used to predict the FC of road vehicles based on on-board data in [143]. Another application is in [144], where the estimations of FC is used for heavy vehicles by combining the data from GPS, road, vehicle, and weather.

The studies in [145] showed that RF technique produces a more accurate prediction with Mean-squared-error (MSE) equal to 0.001 (equivalent to a root mean square error (RMSE) of 0.04) compared to both gradient boosting and NNs in FC modelling of a long-distance public bus, given all available parameters as a time series. In [146], RF slightly outperforms SVR and NNs in FC prediction for trucks. This also includes extra-vehicle features, together with road condition data such as road geometry and the condition of the road infrastructure, derived from fleet managers and road agency databases which affect significantly the fuel economy.

2.4.3 Support Vector Machine for Regression (SVR)

SVR is a version of the support vector machine (SVM) for regression [147]. Fig. 12 illustrate an example of an SVR model. SVM [148] implements the principle of structural (instead of empirical) risk minimization for excellent generalization ability in the situation of a small training sample [149]. By using the "kernel trick", SVMs can change a nonlinear learning problem into a linear separation one, for reducing the algorithm complexity. It is an effective technique in the modelling of complex functional correlations with complex mathematics behind [150].

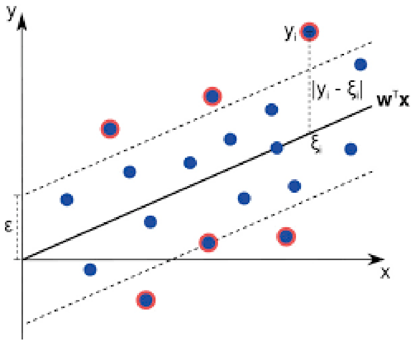


Fig. 12. Illustration of support vector regression (SVR) function [151].

2.4.4 Artificial Neural Networks (ANNs)

With inspiration in the biological neural networks (NNs) of the human brain (Fig. 13) [152], ANNs exhibits features such as the ability to learn complex patterns of data and generalize learned information. It is mostly used to estimate or approximate complex nonlinear relationships between explanatory and response variables [153][154]. Those techniques are composed of several computational elements that interact through connections with different weights. This structure supports parallelization and the ability for adaptive learning - solve problems with collective processing, self-organization and fault tolerance.

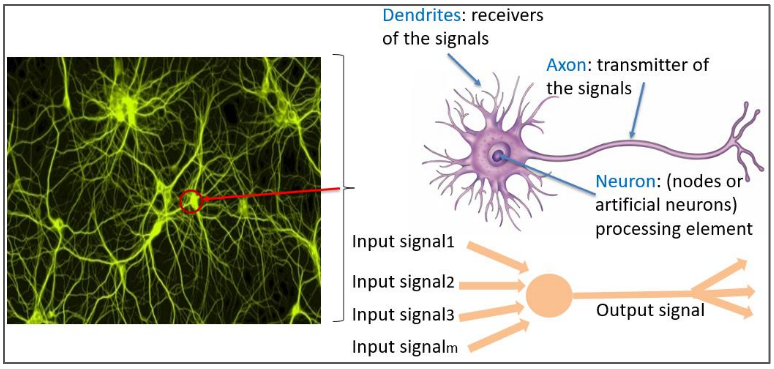


Fig. 13. The similarity between ANN and human neural network.

2.5 Summary and Discussion

This chapter presents some crucial relevant concepts, a survey for the related previous studies, findings and critical analyses to depart my study. I began by showing the importance of driving data analysis, where contextual information from vehicles is fundamental to better understand traffic patterns, driver's behaviour and mobility patterns in a city. Car data can be gained via the CAN bus, that can be queried through the automobile sensing and communication protocols, OBD. Later on, I reviewed the eco-driving technologies as a promising and effective approach in learning and training towards keeping drivers aware of fuel economy. Yet, some of the eco-driving rules, coaching studies and its training types have been discussed together with controversial issues in the definition of this approach. The engaging traits of electronic games are being used to increase engagement and participation in several programs while accelerating learning. Hence, the importance of providing feedback to drivers through gamification's motivation – that I spot on – towards keeping drivers aware of fuel-efficiency continuously has been discussed.

To the best of my knowledge, the technological details of the manufacturer-specific solutions (e.g., Honda's Eco Assist mentioned in 2.3) towards efficient driving are not disclosed in the literature. Following this industrial direction, I intend to support the development of new gameful solutions for fuel efficiency improvement, through a tool available to all the drivers, by using the standard OBD-II vehicular interface and exploiting ML and AI algorithms for extracting RT information about the drive and to support coaching.

Aiming at vehicular diagnosis, the standard OBD-II does not include some effective inputs, such as acceleration or brake pedal pressure or engaged gear. This limitation might be overcome by carefully including synchronized signals from other devices, such as cellular phones, that include inertial sensors (accelerometer, gyroscope, magnetometer) (e.g., [32]). Also, there is no unique measure nor definition, nor a method for driving evaluation, where the past studies have used different methods with different inputs toward driving classification.

As discussed in 2.2.5, there is no specific definition for eco-driving. It varies widely from source to another [57]. Safety should be considered carefully in eco-driving gaming applications, since the fact that unsafe fuel-saving comportments will be reinforced any time fuel-saving in general [57], e.g., letting the vehicle to move not under the driver's control with coasting function of hypermiling while driving (check 2.2.5). Also, eco-driving systems have to be non-intrusive and not highly demanding user attention for driver's distraction and irritation avoidance.

3 In-vehicle Data Source Processing Unit and Primary Experiments

Contextual information from vehicles is fundamental to better understand driver's behaviour, traffic and mobility patterns. In order to proceed with the analysis and to define and measure the accuracy of the targeted models, annotated public dataset containing a significant number of car sensors, recorded by different drivers, and representing several driving styles, is fundamental. The present chapter provides the driving data for the modelling and the simulation stages for this study. Therefore, sections 3.1 introduces the enviroCar open driving data platform that this project fed on. Section 3.2 explains the enviroCar data design that is crucial in the creation of a local database to store some of the requested data (detailed in 3.3). The data pre-processing is summarized in 3.4. Some primary experiments are demonstrated in section 3.5.

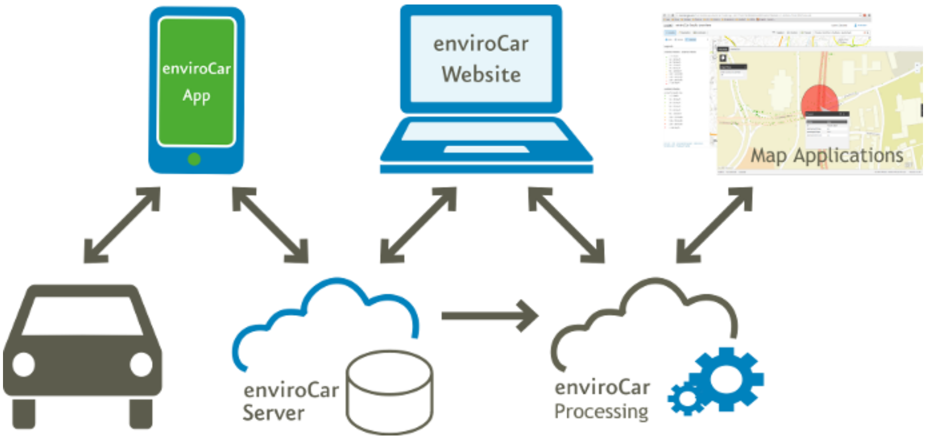


Fig. 14. Overview of enviroCar system design [26].

3.1 EnviroCar Driving Open Data

For the experiments of this research, all used data comes from naturalistic drives, where I relied on the enviroCar platform. It is a community-based open data collection platform for gathering pseudonymized naturalistic driving vehicular sensor data (cars are identified by ID numbers only) and producing environmental information (FC and CO2 emissions) [26].

On the hardware side, this community uses standard Bluetooth OBD-II adapters, that read information from the onboard car sensors through the vehicle CAN bus (section B) upon the PIDs (2.1.3). This information is sampled at regular time intervals (most of the tracks I used are recorded at a 5 seconds sampling time) by an Android smartphone app, and delivered to the enviroCar server, together with GPS information for spatial-temporal analysis. Any citizen who owns an Android mobile phone can download the enviroCar mobile application to contribute by publishing his/her driving tracks as anonymized open data. This feeds the open science community and then encourages external scientists and stakeholders to contribute by analyzing and enhancing concepts in the field of mobility and environment by providing access to their collected measurements on the Web for subsequent analysis. Green drive parameters, such as FC and CO2 emissions estimation, are computed post-hoc and added on their server. An overview of the system design, consisting of the enviroCar Android mobile app and the enviroCar server is depicted in Fig. 14.

3.2 EnviroCar Data Model

The design of enviroCar data model is depicted in Fig. 15, where

1. “Measurement” (feature) class is the centre of the model. This concept comprises multiple properties, such as ID, location, and a timestamp for each measurement.
2. “Phenomenon” entity provides the values with the associated unit for each recorded sensor.
3. “Track” class represents the length of the trip.
4. “Sensor” entity presents information about the car characteristic. It comprises 2 types “Car” or “FeatureCollection” in case the sensor object refers to a car object.

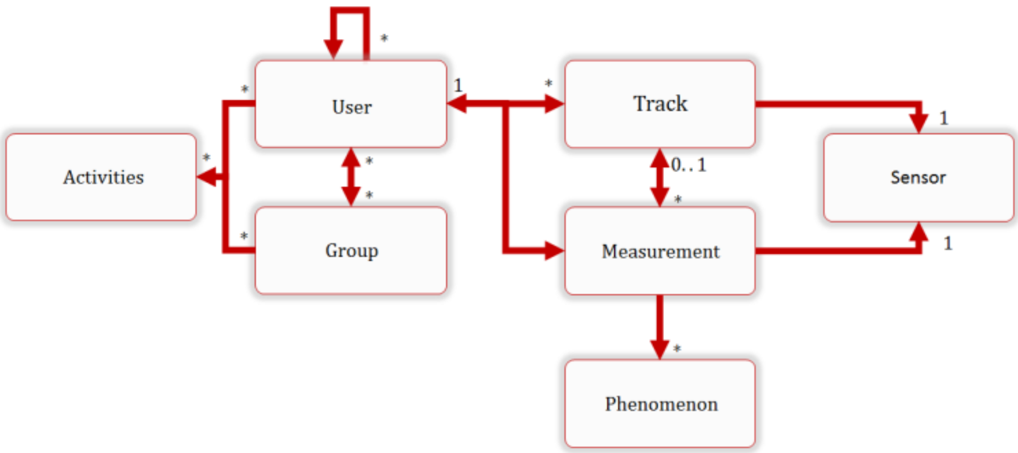


Fig. 15. EnviroCar data model design [26].

EnviroCar supports a subset of the OBD-II available parameters, which are standardized for the OBD-II interface – the ones that are most reliably supported by all car manufacturers. The data providers can be divided into four groups;

(1) Information relevant to car characteristics, including (a) car manufacturer (e.g., Mitsubishi, BMW, Volkswagen), (b) car model (e.g., Fabia and oktavia), (c) car construction year, (d) fuel type (gasoline, diesel) and (e) engine displacement (e.g., 2200, 1600).

(2) Information collected from the car internal sensors through the OBD-II system;

1. Car speed measured in “km/h”: provides the actual speed of the car, that is shown by the odometer. When no value from the speed sensor, I have relied on GPS speed.
2. Engine speed measured in “u/min”: denotes the number of engine revolutions per minute (RPM).

3. Calculated engine load ranging from 0% to 100%: illustrates the power that the outside world takes away from the engine. It measures how much air and fuel are sucking into the engine.
4. Throttle position sensor (TPS) or the throttle pedal and it is also known as 'accelerator', ranging from 0% to 100%: is so named because it regulates the air and fuel intake into the engine, making it run slower or faster – the more the throttle pedal is depressed, the more fuel and air will be supplied to the engine and ignited.
5. Intake air temp measured in “c”: provides the temperature of the air entering inside the cylinders.
6. Air flow rate or mass air flow (MAF) measured in “l/s”: measures the amount of air that is flowing into the engine. The engine control unit uses it to determine how much fuel has to be inserted into the cylinders.
7. Intake manifold absolute pressure (MAP) measured in “kPa”: is used in an internal combustion engine's electronic control system for the estimation of FC for diesel cars (which is not involved in this study).
8. Oxygen sensors; (a) current measured in “A”, (b) current ER, measured in “ratio”, (c) voltage, measured in “V” and voltage ER, measured in “ratio”: is used for FC’ estimation for diesel engines, which is not involved in this study.
9. Fuel trim sensor measured in “%”: shows the percentage of change in fuel over time. It is used to balance the air and fuel.
10. Fuel system used to control the combination of parts needed to carry fuel into and out of the engine; (a) loop, expressed as “boolean number” and (b) Status Code, expressed in “category” (open loop or closed-loop).

(3) Information obtained from embedded sensors in the driver’s smartphone; (a) timestamp for recording the date and the time of every measurement of a track, and the (b) GPS parameters:

- GPS speed measured in “km/h”: represents the current speed of the car.
- GPS accuracy measured in “%”: involves the current accuracy of the signals broadcasted in space by GPS satellites. However, the reception depends on additional factors, including satellite geometry, signal blockage, atmospheric conditions, and receiver design features/quality [155].
- GPS three-dimensional location includes; (a) latitude, (b) longitude, and (c) altitude.
- GPS Bearing measured in “Deg”: provides the direction to take to move toward a destination.
- Dilution of Precision (DOP) measured in “precision” – is a factor that multiplies the uncertainty associated with user equivalent range errors, represented in the following user local coordinates:
    - Horizontal Dilution of Precision (HDOP) for local horizontal.
    - Vertical Dilution of Precision (VDOP) for local vertical.

- Position Dilution of Precision (PDOP) for the three-dimensional position, a combination of the two previous components [156].

(4) Estimated variables by enviroCar and provided to the public; FC, Co_2 emissions, calculated MAF and length for each track.

1. Fuel consumption (FC), measured in “l/h”: The community relies on the MAF sensor for its estimation. Consistently, the present analysis shows a Pearson Product Moment Correlation (PPMC) value equal to 1 between FC and MAF (Fig. 19). Tests conducted in [157] show that the MAF is the best candidate for estimating FC from OBD data. The MAF measures the amount of air that flows into the engine, maintaining a fuel-air ratio of the automobile. It gauges the volume of air entering to the vehicle's fuel injection engine and sends that information to the engine control unit (ECU) to correctly balance and deliver the correct amount of fuel to the engine. Then, the ECU determines the amount of gas that fuel injectors need to send to each of the cylinders. Consequently, the MAF plays a vital role and it is particularly important to the engine controller, where any fault or defect of this sensor can affect the performance of the emissions control system of the engine [158][159][160].

   EnviroCar estimates FC for gasoline vehicles with the formula in Eq. (1) [161]. MAF is measured in (grams/second), AFR is the Air Fuel Ratio (Mass Ratio of Air to fuel), which is 14.7 for gasoline. MAF over AFR is thus in (grams/second). EnviroCar provides the FC in (litres/h), where gasoline has a 745 (grams/litre) density (see eq. (2)). The combustion is complete for a ratio of 14.7 kg of air per 1 kg of gasoline [26].

$$\text{Fuel consumption}(\mathrm{g}/\mathrm{s}) = \frac{Fuel\ weight}{\mathrm{s}}(g/s) = \frac{MAF(\mathrm{g/s})}{\mathrm{AFR}} \tag{1}$$

$$\text{Fuel consumption}(\mathrm{l}/h) = \frac{\frac{MAF(\mathrm{g/s})}{\mathrm{AFR}}}{Density\ of\ fuel\ (g/l)}(l/s) * 3600(l/h) \tag{2}$$

   On the other side, although the MAF sensor is mandatory in the OBD-II standard, it is not supported by some vehicle types. Thus, enviroCar estimates it by utilizing other parameters, namely temperature, air pressure, and engine speed [26][162]. In this work, the focus is on gasoline engines, as the FC’ estimation by this community provides the best accuracy for a gasoline engine rather than diesel engine [163].

2. CO_2 emissions of the engine, measured in “kg/h”: they are not influenced by filters or catalysts, which makes its estimations easier than other pollutants [164]. It is derived as in the eq. (3), based on the current FC, since it is in a linear relationship with the fuel mass that is combusted in the

engine (validated in Fig. 21). In the case of the gasoline engine, the combusting of 1 litre results in 2.35 kg of CO_2.

$$Co2\_gasoline(kg/h) = \frac{(\text{MAF}/14.7)}{745} * 2.35 \tag{3}$$

3. Calculated MAF in "g/s": this sensor is not supported by all car types although it is mandatory in the OBD-II standard. When OBD adapter delivers no data response for MAF, the community estimates MAF by utilizing other readable parameters, namely temperature, air pressure, and engine speed [162].
4. Track length measured in "km": the travelled distance in kilometres (or the distance from the startup point to the final point) of an individual track.

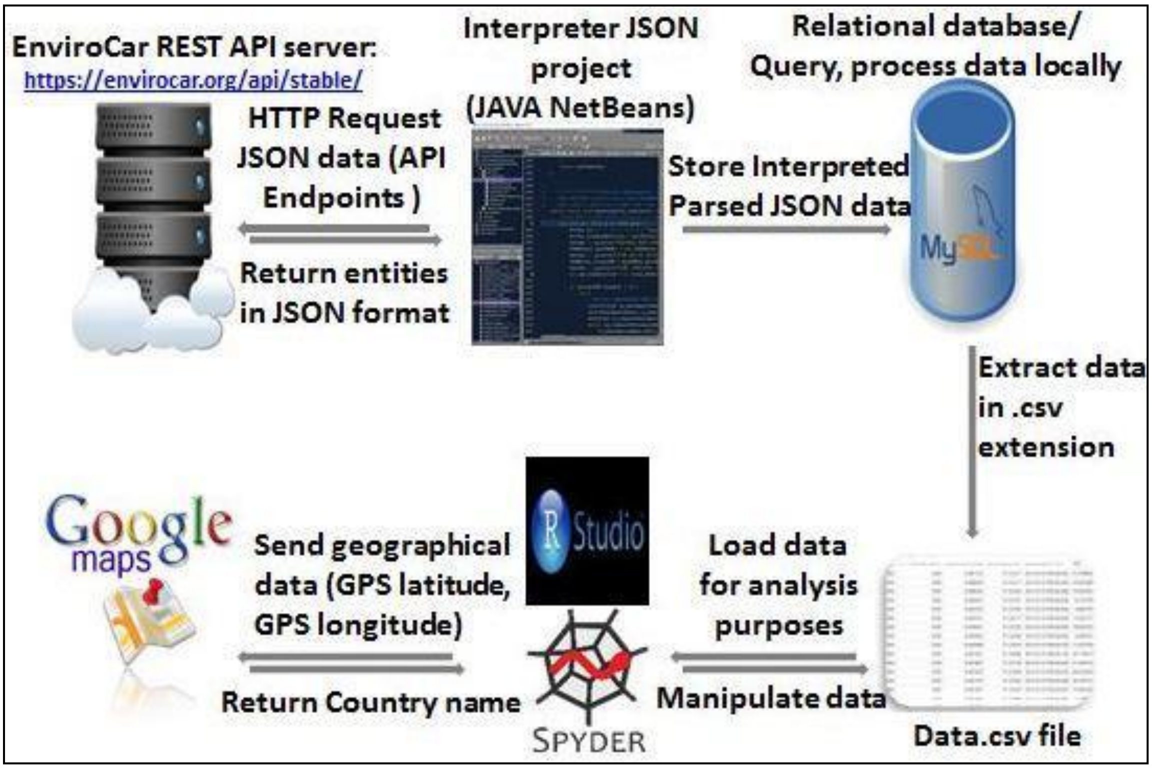


Fig. 16. Data system experimental architecture.

3.3 Data Acquisition

To build the dataset, I developed a software system written in Java programming language (using NetBeans integrated development environment (IDE) for Java), that requests the anonymized track data through a JavaScript Object Notation (JSON) interface, using enviroCar REST API via HTTP requests (Fig. 16). To protect their server performance, the community follows the pagination process, limiting the request to retrieving 100 entities in maximum at once. The project analyses and then parses the requested JSON data to stores them later into a relational local database (using MySQL workbench tools) designed and created based upon the enviroCar data model (Fig. 15) for querying purposes (Fig. 16) such as filtering the created local database based on specific car characteristics (e.g., data model, manufacturer and construction year).

For the analyses, 8726 different gasoline tracks have been considered, with 983291 complete record measurements for gasoline engines. They were recorded mostly in Germany in the period 2012-01-01 – 2016-06-15. For retracing the journey of each car, an ID number identifies each of the instantaneous features, besides the timestamp extracted from enviroCar smartphone application.

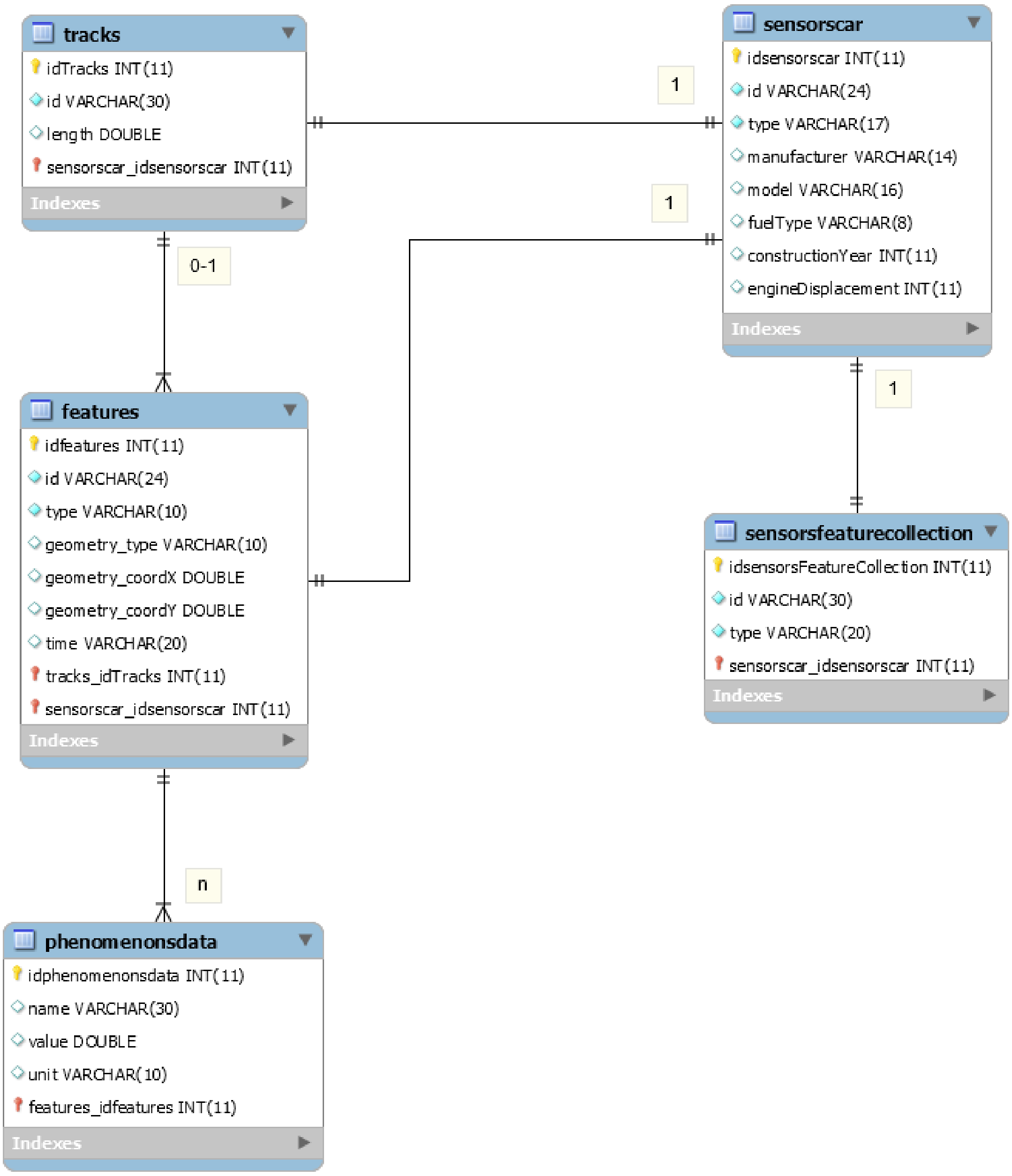


Fig. 17. MySQL workbench extended entity-relationship model diagram.

3.4 Data Pre-processing

The data pre-processing involves the definition of data' fields to use and data manipulation. The data manipulation covers fixing obvious missing or erroneous car characteristics' inputs by the drivers who have contributed by their track data to “enviroCar” platform. For instance, it covers unifying the way of writing the two fields ‘car manufacturer’ and ‘car model’ e.g., “Volkswagen”, “VW Golf” and “VW” are unified to be under the manufacturer “Volkswagen”. Furthermore, the field “engine displacement” has been requested by the car manufacturer, model and construction year when there is no value.

For departing my data analysis, the requested locally data needs to be in the form of the data frame, so data should be merged into one data frame (Fig. 18). I queried MySQL to extract the data in different datasets (in the form of excel files) to merge all of them into one unique excel file. Merging the datasets was easy for some datasets via the joining process. While it was a bit challenging for some. This is the case while merging the “phenomenon” data that represents the sensors (Fig. 15), where each feature has many sensor data fields. Thus I requested separately each sensor in the data and then I merged all the CSV files of the sensors to my data based on “feature ID”.

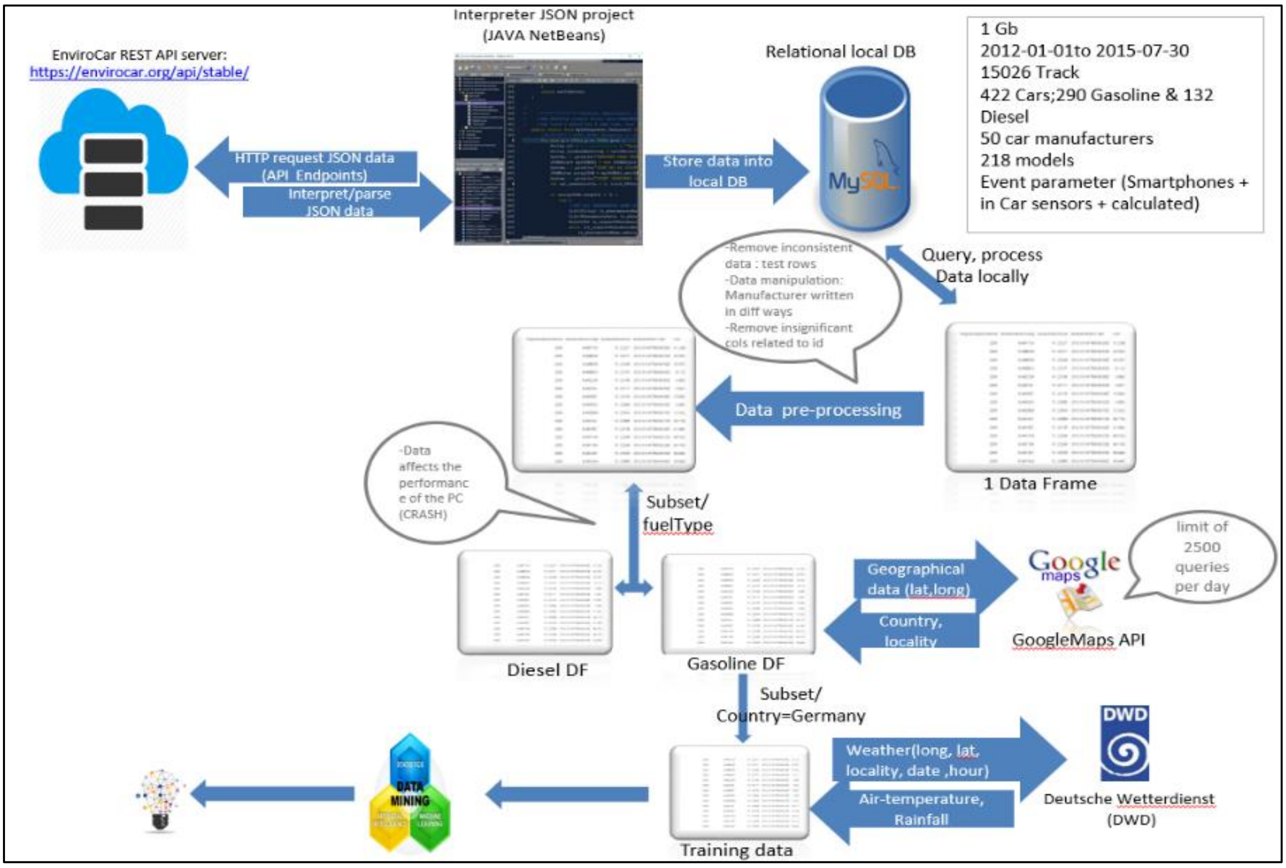


Fig. 18. Data pre-processing.

After the merging process, I reduced the size of the data. Some very short test trips were excluded (e.g., relevant to test, tester, test2). Yet, I excluded some unnecessary columns (were necessary for the merging procedure such as the ones related to ID and FK) resulted from the data structure of MySQL (Fig. 17) respecting enviroCar data model discussed in 3.2. Likewise, the exclusion involves the fields “measurementGeomType” (referring for geometry_type which is the point for the entire data that I requested), “measurementType” (all are Feature) and “VehicleType” (all are a car). Likewise, I excluded the two columns referring to both sensors “Fuel System Loop” and “Fuel System Status Code” since they have no values in all my local data set. The MAF and the calculated MAF (in case OBD adapter delivers no result for MAF) have been merged in one column wherein there is no readable MAF from OBD-II, I relied on the calculated one by “enviroCar” [162].

Secondly, some external data was added to the data frame. In order to recognize the country where a track was recorded, I used the Google Maps API [165], provided by the “*ggmap*” R library [166], for the process of back (reverse) coding for a point location (GPS latitude, GPS longitude) into a readable address (country, locality, and route) [167]. This helps in filtering the trips for a specific location. Yet, these additional data are useful for requesting weather information.

Since weather affects the fuel economy level (1.2.3), I have requested historical hourly weather data (Fig. 18) to consider in my data analysis. Considering that most of the tracks were recorded in Germany, the weather data was requested from the German Weather Service (Deutscher Wetterdienst, DWD) through the FTP server online at ftp://ftp-cdc.dwd.de/pub/CDC/observations_germany/climate. I focused on two parameters, the air temperature (historical hourly station observations of 2m air temperature in ° C) and the precipitation (historical hourly station observations of precipitation for (i) hourly precipitation height in mm, (ii) Yes/No precipitation and (iii) form of precipitation). I relied on the “*rdwd*“ R package for climate data from the German Weather Service, that contains code to select, download and read weather data from stations across Germany [cran.r-project.org/web/packages/rdwd/vignettes/rdwd.html]. The process consists on searching the nearby weather station (s) (www.rdocumentation.org/packages/rdwd/versions/0.8.0/topics/nearbyStations), which could be determined based on each feature geographical data (GPS latitude and GPS longitude) and the state name requested before with Google Maps API. The information of the stations is downloaded later locally as a zip file to be requested using two parameters; the date and the hour for each measurement in a track. Note that the process may be switched to the next nearby station if no records have been found for the feature date and hour location in a station.

3.5 Primary Experiments

As stated earlier, the focus here is on track data for gasoline engine. In order to determine which sensors (individually or combined) provides valuable information about the vehicle, the driving attitude and affect fuel economy simultaneously, I first need to characterize their readings in previously known contexts. For an initial analysis, I considered 12 out of the monitored variables. Some are directly readings from the vehicle's sensors, others are calculations based on data collected from the car and others are measured using the smartphone's sensors (refer to 3.2). The PPMC depicted in Fig. 19, shows the correlations in a symmetric matrix between some OBD-II variables and the estimated FC by the community (also the estimated Co2) in the local database, that represent both lines and columns of the matrix.

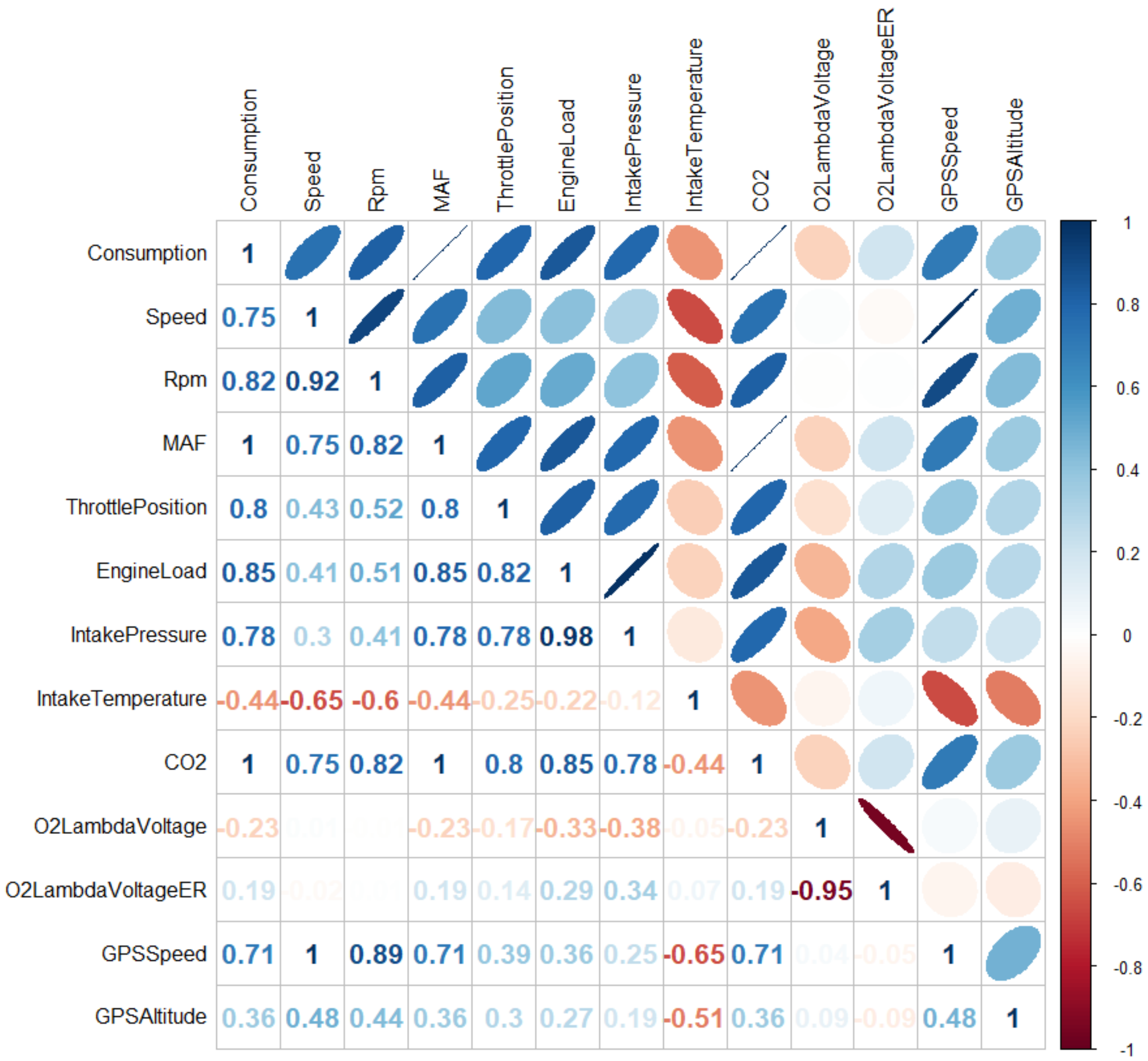


Fig. 19. Correlation between enviroCar estimated FC and car variables [ellipse shapes and colours are explained in the text section 3.5].

The guidelines to interpret the explicit values of PPMC are provided in Table 2. A positive sign refers to a positive correlation, while a negative sign indicates a negative correlation. The closer the value between an input and the outcome FC to 1 or -1, the more the variable is likely to be predictive. Since the correlation matrix is symmetric, on one side it shows the explicit values of the correlation and on the other side, the same value is visually shown as the ellipse that is expected from a bivariate distribution with the same correlation value (high correlation values between 0.5 to 1.0 or -0.5 to -1.0; a medium correlation between 0.3 to 0.5 or -0.3 to -0.5; a low correlation between 0.1 to 0.3 or -0.1 to -0.3; No correlation when 0). Thus, visually, ellipses close to straight lines represent two tightly linked sensors, which can be directly or inversely correlated, depending on the line direction. On the other hand, variables with a small relationship are represented by an almost invisible circle, due to the colour scale. For instance, there is a high positive correlation between FC and RPM with PPMC value is equal to 0.82. An example of an inversely proportional relationship is between the intake temperature and car speed (PPMC=-0.65).

Table 2. Guidelines to interpret pearson correlation coefficient.

Linear Association Strength	+ Association, colour	- Association, colour
Small	0.1 to 0.3, light blue	-0.1 to -0.3, light coral
Medium	0.3to 0.5, mid blue	-0.3 to -0.5, mid coral
Large	0.5 to 1, dark blue	-0.5 to -1, dark red
No correlation	0, white	0, white

Fig. 20 depicts the plot pairwise scatter plots which help in teasing out the relationships between the variables quickly. For instance, FC increases with the increase of car speed (OBD-II speed, GPS speed), RPM, TPS and intake pressure.

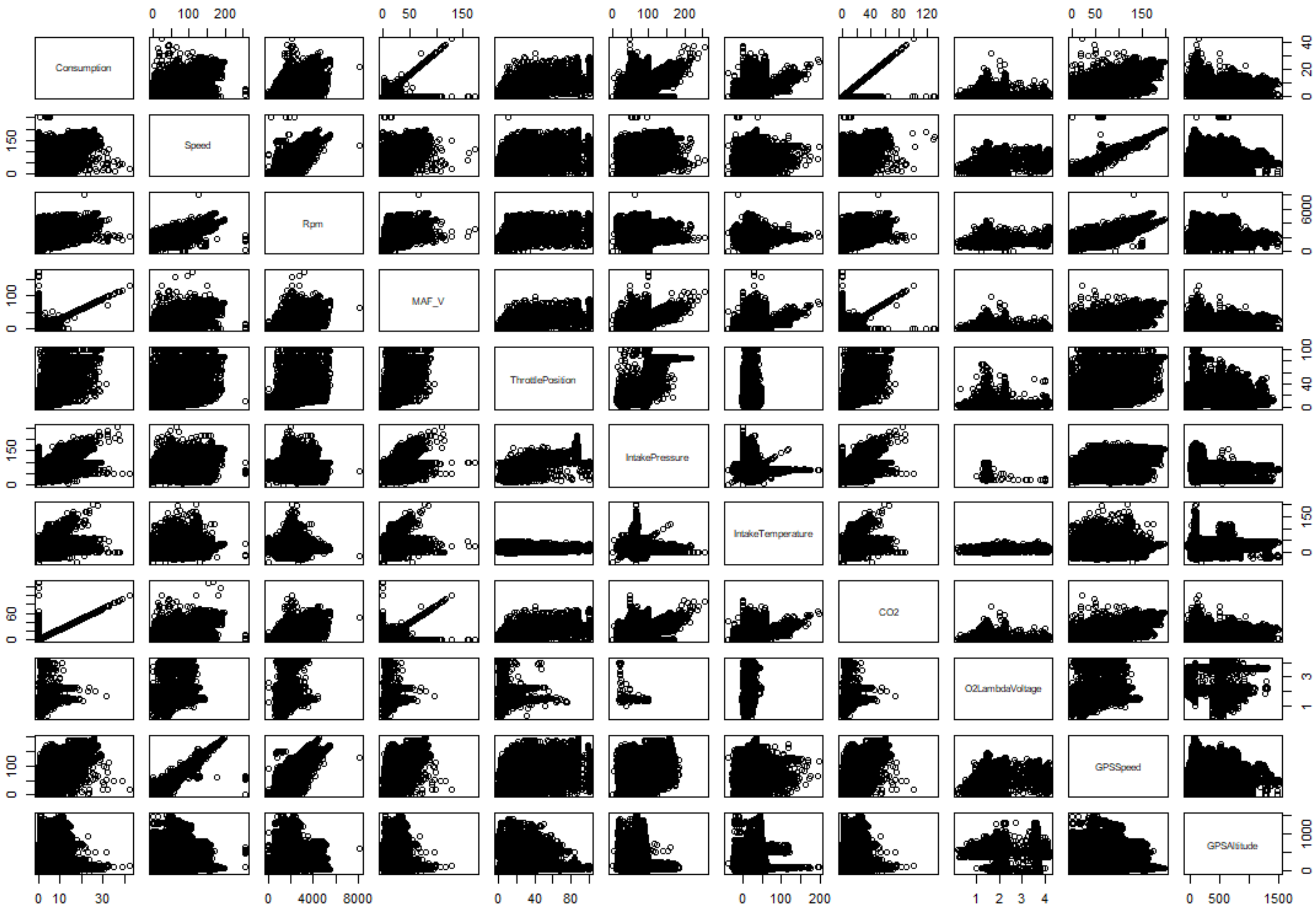


Fig. 20. Pairwise scatter plots.

3.5.1 Fuel Consumption - CO_2 Emissions Correlation

The scatterplot for all the local gasoline tracks (Fig. 21), the Co2 emissions dependent directly on FC.

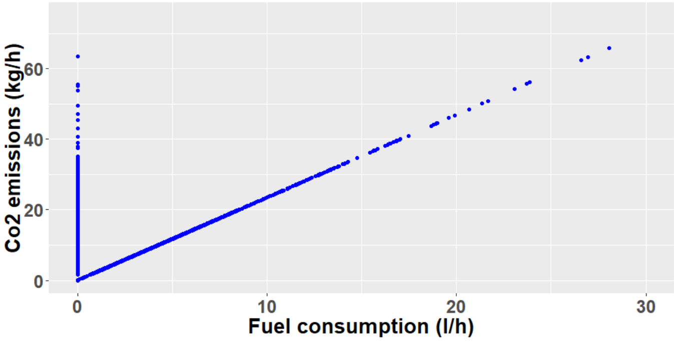


Fig. 21. FC – Co2 emissions correlation (local tracks).

3.5.2 OBD-II Vehicle's Speedometer overlaps with Mobile GPS Speed.

There is a debate concerning whether the recorded speed by the vehicle speedometer or GPS vehicle tracking device is more accurate when analyzing data. However, some factors affect the accuracy of both sensors. The accuracy of GPS may vary slightly as the vehicle travels from areas with a clear sky to those without. For example, areas that have heavily three-lined streets, tunnels and covered parking lots can an issue when it comes to GPS inaccuracy.

On the contrary, it is much more difficult to prove how some factors affect the precision of the speedometer. The most notable factor is the differences in wheel size due to wear (e.g., a reduction of just one inch from the stock tyre size would result in the speedometer reading approximately 5 % faster), tire pressure and temperature [168][169].

GPS velocity data is noisy comparing to the velocity read from the OBD-II system [39]. It is obvious from the scatterplot in Fig. 22, that the GPS speed and the OBD-II speed are two highly correlated sensors with some noises. I considered in this study the velocity read from the OBD-II system.

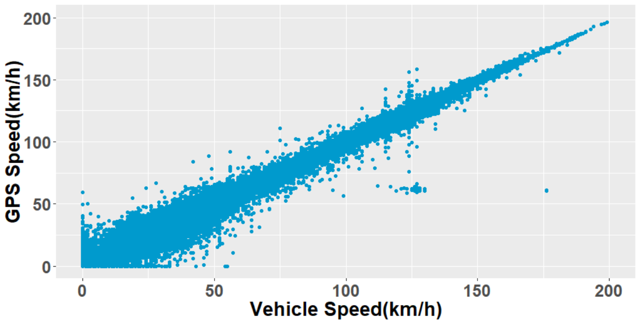


Fig. 22. Vehicle speedometer (OBD-II speed) - GPS speed correlation.

3.5.3 Vehicle Speed - RPM Correlation

Fig. 23 depicts the plotting data for RPM and vehicle speed for a track recorded by a manual car (gasoline Opel Vectra C Caravan 2004). There are clear groupings of points that share a stronger relationship that is equivalent to the gear ratios of the vehicle. The gears are shown by the six different lines.

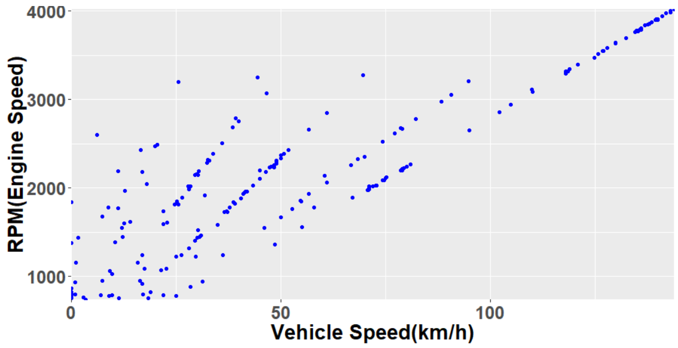


Fig. 23. Vehicle speed - RPM correlation (one track).

3.5.4 RPM – Fuel Consumption Correlation

Fig. 24 illustrates the correlation between RPM and FC for to all local gasoline tracks. Engine revolutions have a very significant influence on FC – the higher the RPM, the more the fuel is consumed.

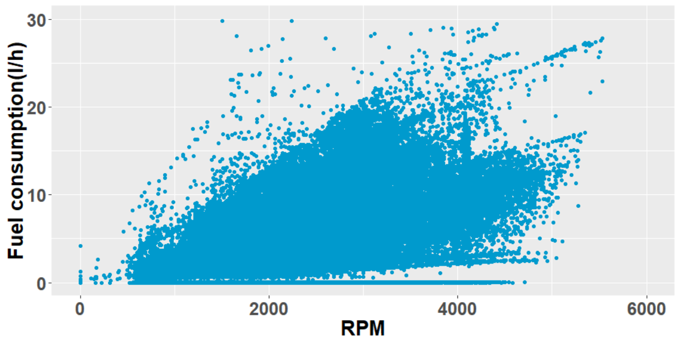


Fig. 24. RPM - FC correlation.

3.5.5 Vehicle Speed - Fuel Consumption Correlation

Driving too fast requires fuel burning (Fig. 25). Concerning unstable speed and idling for a longer time (RPM has value while vehicle speed=0) effects, they need more investigation in the function of time.

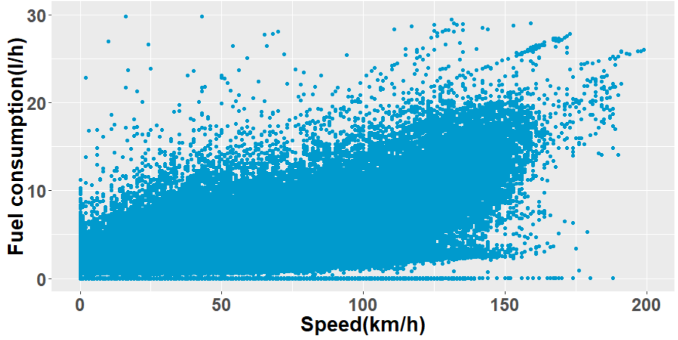


Fig. 25. Vehicle speed - FC correlation.

3.5.6 Engine Load – Fuel Consumption Correlation

The more the engine is loaded, the more fuel burns (Fig. 26).

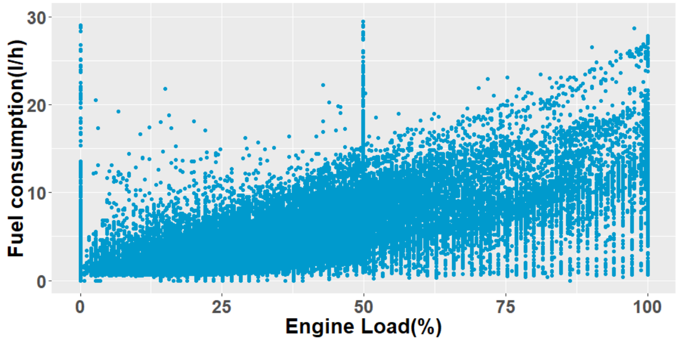


Fig. 26. Engine load - FC correlation (Local tracks).

3.5.7 Engine Displacement - Fuel Consumption Correlation

The power of a car engine increases with the increase of the number of engine's displacement. Hence, FC goes up with an increase in the number of engine displacement. This is clear at the beginning of the scatterplot in Fig. 27. The plot shows that the consumption of fuel can be controlled even while having a high value of engine displacement.

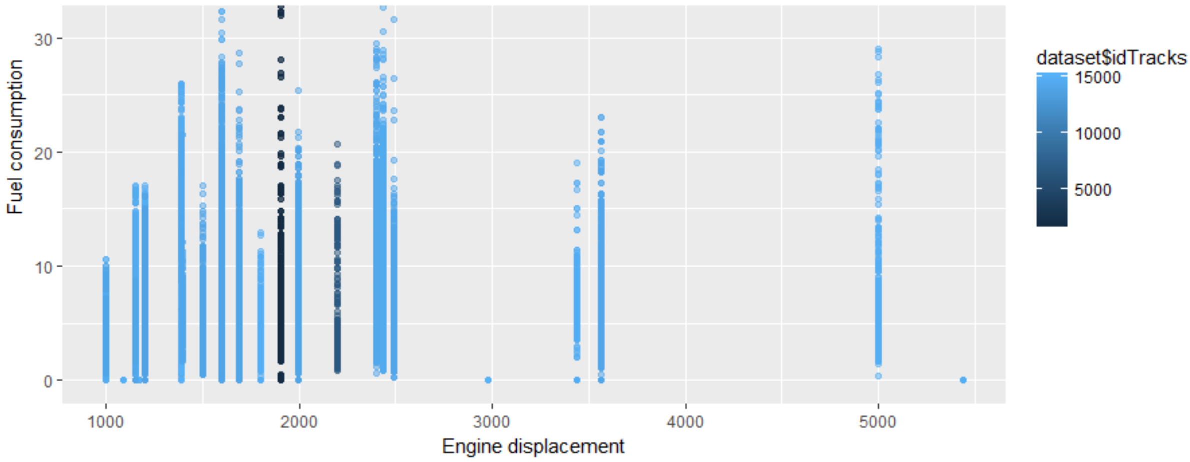


Fig. 27. Engine displacement - FC correlation.

3.5.8 Throttle Position (TPS) - Fuel Consumption Correlation

TPS – also known as acceleration or gas pedal position (angle) – controls fuel and air supplying to the engine. It is usually located on the butterfly spindle/shaft so that it can directly monitor the position of the throttle. When a driver accelerates, the butterfly opens a little or a lot to allow the entrance of air into the intake manifold, depending on how much the driver is accelerating. The more the butterfly is open, the more fuel that enters the engine; the engine then calculates how fuel is consumed based upon how open is the throttle. The increase in air and fuel in the cylinders of the engine as the gas pedal is pressed increases FC (Fig. 28).

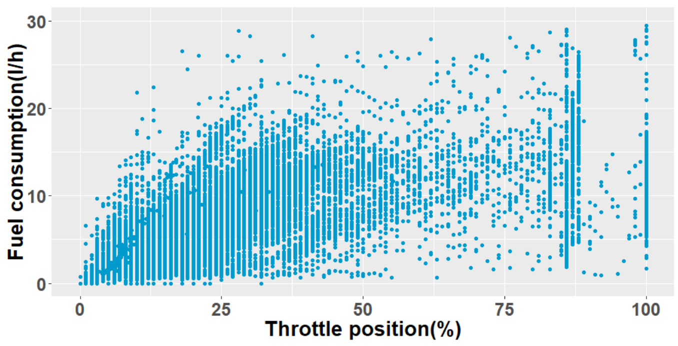


Fig. 28. TPS-FC correlation.

3.5.9 Throttle Position (TPS) – Engine Load Correlation

Engine accommodates more load when its throttle is open wider, while less power when its throttle is less open. As air and fuel enter into the cylinders of a car engine once the throttle is open slightly, as the engine load increases (Fig. 29).

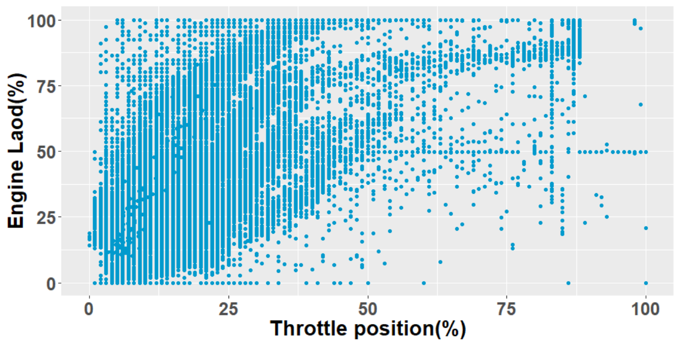


Fig. 29. TPS - Engine Load correlation.

3.5.10 Car Construction Year – Fuel Consumption Correlation

Fig. 30 depicts the effect of construction years on FC for trips relevant cars with to gasoline engines, constructed between 2000 and 2014. Recent cars are more fuel-efficient than the old ones, the drop in FC level in the 2000-2005 period. However, upon the local data, some rises in FC levels occur, for example in 2006 and 2013. This scatterplot validates that even though purchasing fuel-efficient vehicles increases the fuel economy, other influences may affect FC (e.g., driving behaviour).

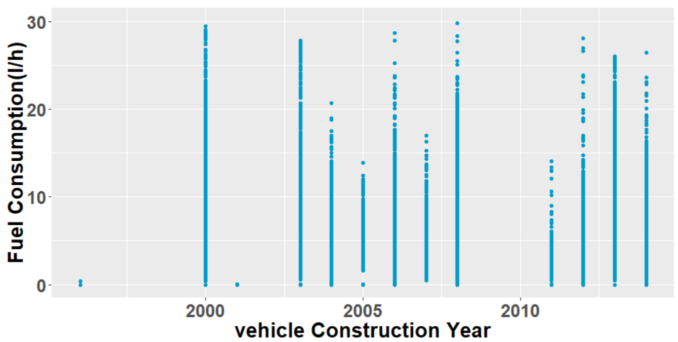
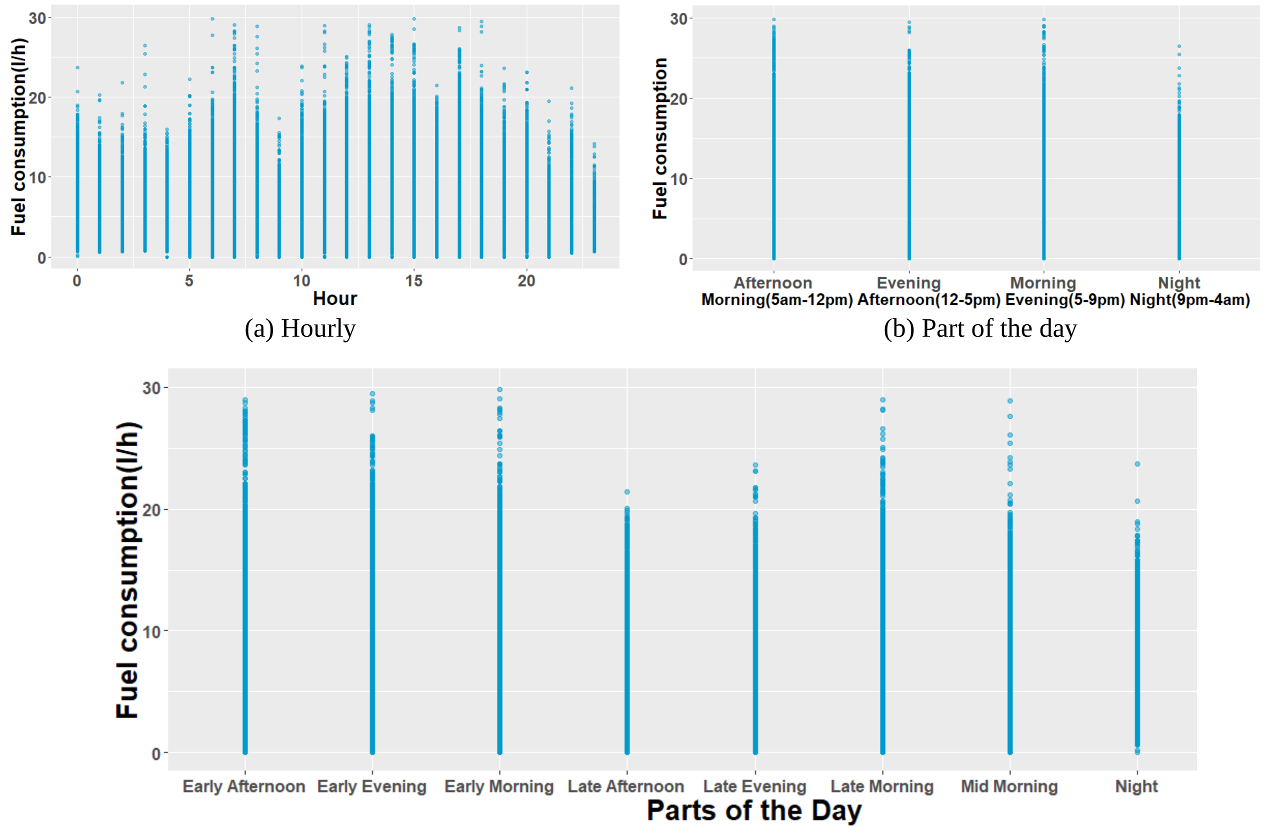


Fig. 30. Car construction year - FC correlation.

3.5.11 Hourly Time - Fuel Consumption Correlation

Fig. 31 (a) depicts the influence of time (extracted from the timestamp) on FC. The time is divided into four subclasses: Morning (5 am to 12 pm); Afternoon (12-5 pm); Evening (5-9 pm) and Night (9 pm to 4 am) in Fig. 31 (b). Yet in Fig. 31 (c), the FC correlation is illustrated with eight subclasses associated to DTP (day time period); Early morning (5-8 am); Mid-morning (8-11 am); Late morning (11 am to 12 pm); Early afternoon (1 to 3pm); Late afternoon (4 to 5pm); Evening 5pm to 9pm; Early evening (5-7 pm); Late evening (7-9pm); and Night (9pm to 4am).



(a) Hourly (b) Part of the day (c) Part of the day detailed

Fig. 31. Time influence on FC.

The lowest FC level is at night, which is clear from the beginning (00 to 4 am) and the end of the graph (21-23 pm) in Fig. 31 (a). The FC level goes up in the early morning (5 to 8 am) which is the rush time (work, school etc.). There is a sudden drop in FC in the mid-morning (8 to 11 am), where people reach their destinations (e.g., employees reach their workplaces, children reach their schools). The trend increases back from the late morning (11 am to 12 pm) until the early afternoon (13-15 pm). This is probably the reason for finishing some works and returning from schools. Between 16 pm and 17 pm (Late afternoon), the level decreases to went up at 17 (a rush hour to leave most of the full-time jobs). A decrease then happens to be stable in the evening (17-21 pm).

The consumption of fuel is the highest in the morning time slot, followed by afternoon, evening and by the night respectively (Fig. 31 (b)). In Fig. 31 (c), early afternoon and early morning have the highest FC values. This is followed by early evening and late morning, then by late evening, mid-morning, late afternoon and night.

3.6 Summary and Discussion

This chapter describes the data acquisition process from enviroCar - a naturalistic driving open data archive. Some of their data has been requested to be locally stored into MySQL database for querying purposes using spatial and temporal filters. From this data, the focus was on gasoline engines, as enviroCar FC estimation provides the best accuracy for a gasoline engine rather than diesel engine [26][162][163]. The enviroCar data sources could be split into four groups upon their information sources: user input for car characteristics, in-car sensors, embedded mobile sensors (GPS and timestamp) and other information obtained by estimation (e.g., FC and Co_2 emissions). This data is used by Mönchengladbach (Germany) administration with the help of a traffic system consultancy (TSC) for including citizens in the evaluation process of different traffic light configurations along major traffic axes [170] and in [171] for creating an ontology design pattern for space-time prisms.

The enviroCar data is not calibrated for a specific car model, recorded in different driving environments, which made the work challenging and robust for real-world conditions, without the need for the driver or player to insert data about the vehicle nor train the system. Moreover, the enviroCar application estimates automatically further information such as FC and CO_2 emissions, which are the main metrics in eco-driving evaluation.

Meanwhile, no real experiments have occurred with the enviroCar Android app, which requires real environment: go a round trip into a car plugged with an OBD-II Bluetooth adapter used to connect a smartphone to the OBD-II system, and then login into the enviroCar website to visualize the experimented tracks.

Car characteristics have an important impact on FC. Buying a fuel-efficient vehicle is the first direct way to save fuel. However, other factors may affect FC. These factors may be a lack of attention to maintenance practices, route selection, and managing vehicle load, in addition to inefficient driving styles. This is might be the reason of the appearance of some fluctuations with my primary data analysis in 3.5, especially in the related scatterplots to car characteristics (engine displacement Fig. 27 and construction year Fig. 30). Other car characteristics have an important impact on FC are not available mainly car weight (reflecting engine weight), horsepower and number of seats in the car (number of passengers increases FC). The increase in the number of the mentioned characteristics, the more fuel is consumed. Nevertheless, this additional information can be requested via the car manufacturer, model and the construction year.

The weekends and holiday have not taken into account in the related scatter plots to time influences, even though I can sense through this information the traffic state, which affects the fuel economy. Some weather data has been requested but I have not involved them yet since their request is time-consuming; it depends on country and locality inputs, which must be requested before using Google Map API (2500 request per day) and this might depend on internet connectivity (sometimes there is return value).

4 Followed Methodologies and Research Progress

Reality-enhanced gaming is an emerging SG genre, that looks beneficial particularly because of its ability to contextualize a game within its real instruction-target environment. A key module for such games is an evaluator, that senses a field user/worker performance and provides consequent input to the SG. I explore this field, focusing on estimating automotive driver performance in terms of FC. This chapter reports on the modelling stage of the proposed approaches for computing driver performance assessment values (specifically, on driving efficiency – the less the fuel is consumed, the more the driving is efficient) usable in third-party RESGs to assist automotive drivers in promoting more fuel-efficient.

For all the proposed models, this research input and simulation data was taken from the enviroCar database as stated in chapter 3. The experimental platform is depicted in Fig. 32. As mentioned before, this project focuses on gasoline engines since the estimation of FC (measured in litres/h, the formula is given in [161]) by enviroCar provides the best accuracy for them [26].

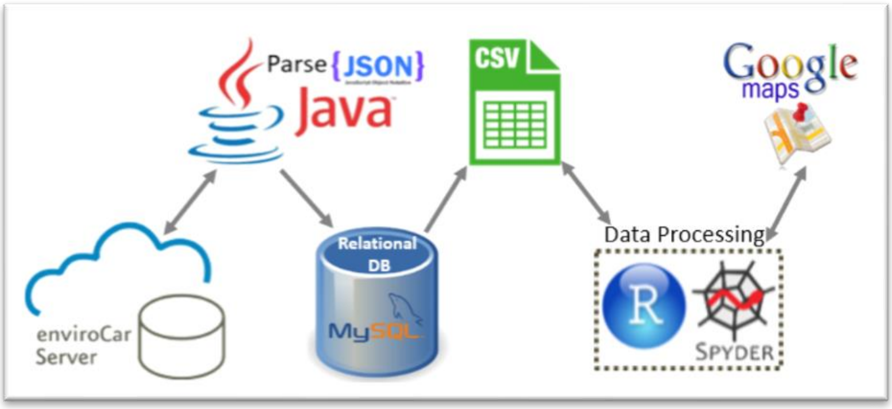


Fig. 32. Experimental system architecture.

To achieve the target of this project (create a module to be pluggable into driving SGs toward eco-friendlier driving style (1.4)), this chapter presents my research' progress which can be organised with four main methodologies:

1. Methodology 1 (4.1): The estimation of FC is challenging work since many influences impact the fuel economy as stated in 1.2. To handle this uncertainty and impression, I departed using the mathematical tool fuzzy logic (FL) which is a form of AI, as a baseline to assess overall feasibility, given its ability to embody expert knowledge and to deal with incomplete availability of information. I relied on the key signals TPS, car speed and RPM.
2. Methodology 2 (4.2): After the implementation of the FL model that is able to provide coaching advice to drivers besides the prediction of FC (4.1), I was interested in exploring if better quantitative FC estimations (with the signals as in 4.1) could be obtained with random forest

(RF) ML technique. The combination of both techniques can supply coaching advice to drivers via the deduced fuzzy rules (4.1), with a more accurate quantitative FC estimation that may be obtained through the RF model (that can be integrated into the game as an energy factor).

3. Methodology 3 (4.3): I work on a new approach for supplying the drivers with direct feedback when no eco-driving events are detectable considering what actions they could perform toward eco-friendlier attitude. Besides the already used key signals (TPS, car speed and RPM), I involved the estimated car jerks (changes in acceleration with respect to time). Significant changes in these, act as events that affect FC to be detected, e.g., detecting drivers overtaking. The thresholds for the sensors' evaluation was set based on the literature review. Again, the quantitative estimations of the FC with the RF module in 4.2, have been improved by involving the OBD-II calculated engine load. On the other hand, considering some driving circumstances which lead to wasting more fuel (e.g., encountering a pothole or stop traffic light timings), it is beneficial to provide an eco-driving score on a scale of 100 for the trip (100 is the best possible score), together with a summary report for the driving pattern. In this process, I involved the achieved fuel-efficiency (the most important metric in eco-driving). The higher the score, the more the driving is efficient compared to other trips relevant to the same car type.
4. Methodology 4 (4.4): I investigate the three ML techniques, support vector machine for regression (SVR), RF and artificial neural networks (ANNs). I involved eleven predictors that affect the FC, including vehicular signals available through the common ODB-II interface for a wide diffusion and a standardized FC module. This covers the case of unreadable or faulty MAF (3.2 (4)). Yet, I relied on the deduced fuzzy rules (4.1) for supplying the drivers with driving advice during driving to avoid fixing a threshold for each one of the inputs as in 4.3.

4.1 Methodology 1: A Fuzzy Module for Fuel Consumption Estimation and Instant Recommendation

I departed with three key vehicular signals, TPS, RPM and car speed. The variables are read from the OBD-II interface, controllable by the drivers and easy to explain for returning feedback to a driver. In a baseline approach, fuzzy logic (FL) has been selected. FL is a promising technique for driving style analysis [85], that gives the possibility of distinguishing between different performance factors, so to give coaching feedback to a driver (refer to 2.4.1). Four models have been studied with all the possible combinations of the chosen variables (TPS and RPM; RPM and speed; TPS and speed; TPS, speed and RPM). The fuzzy models, iteratively defined on the basis of literature expertise and data analysis – can be easily plugged into a reality-enhanced gaming architecture.

TEAM (short for 'Tomorrow's elastic adaptive mobility') is a collaborative industrial project that implements applications including SGs to support flexible collaborative mobility. Its application covers collaborative navigation, car parking, lane merging, public transport optimization, smart intersections and driving corridors [32][172]. The main requirements that I followed in the context of TEAM – concern:

1. RT driver performance assessment with a sampling frequency in the order of a few seconds, and with negligible computation latency (which is compatible with a typical game's timing).
2. Easy access to sensors.
3. Easy to perform profiling and understandable modelling.
4. Distinguishing between different performance factors, so to give coaching feedback to a driver.

The present work leverages the TEAM SG concept architecture. It comprises three layers that distinguish between (1) sensing modules (e.g., the vehicular signals), from (2) game logic (e.g., virtual bank, snake & ladders, races), from (3) user interface (e.g., on a smartphone) [99]. Decoupling these three sub-systems with such architecture allows implementing various types of games, fed (e.g., directly as score or indirectly as an energy factor, or even other game mechanics) by the seamless insertion of sensing modules (e.g., to evaluate the driving style, but also other aspects of driver/user performance). This can also provide different modalities of user interaction.

Hence the proposed architecture can be logically ported to other application domains beyond cars – studying a methodology that processes the signals from the physical sensors on a vehicle to estimate in RT the performance of a user, and then it can be employed as a factor in a SG (e.g., I address techniques for implementing layer 1 in the TEAM architecture). The remainder of this chapter is organized as follows: 4.1.1 presents the feature selection; 4.1.2 provides the modelling process of the proposed fuzzy systems; 4.1.3 illustrates the experiments to assess the models; A summary is drawn in 4.1.4.

4.1.1 Features Selection

I first define a set of vehicular signals that impact fuel economy, that is well understandable and controllable by the driver. Coaching feedback can be supplied to drivers given the fact that fuzzy rules to be applied to such inputs, reflect expert knowledge or at least link estimated outcomes (e.g., FC) with inputs. Then, I develop FL-based estimators, exploiting those sensors on naturalistic driving historical data.

The selected sensors, RPM, TPS and car speed are considered in the five common rules of eco-driving [43][173][174], and they have been recently used to discriminate between different driving styles (e.g., [12], also using FL [132][175]). These variables are also easy to access from the car – directly available from the OBD-II standard interface [45].

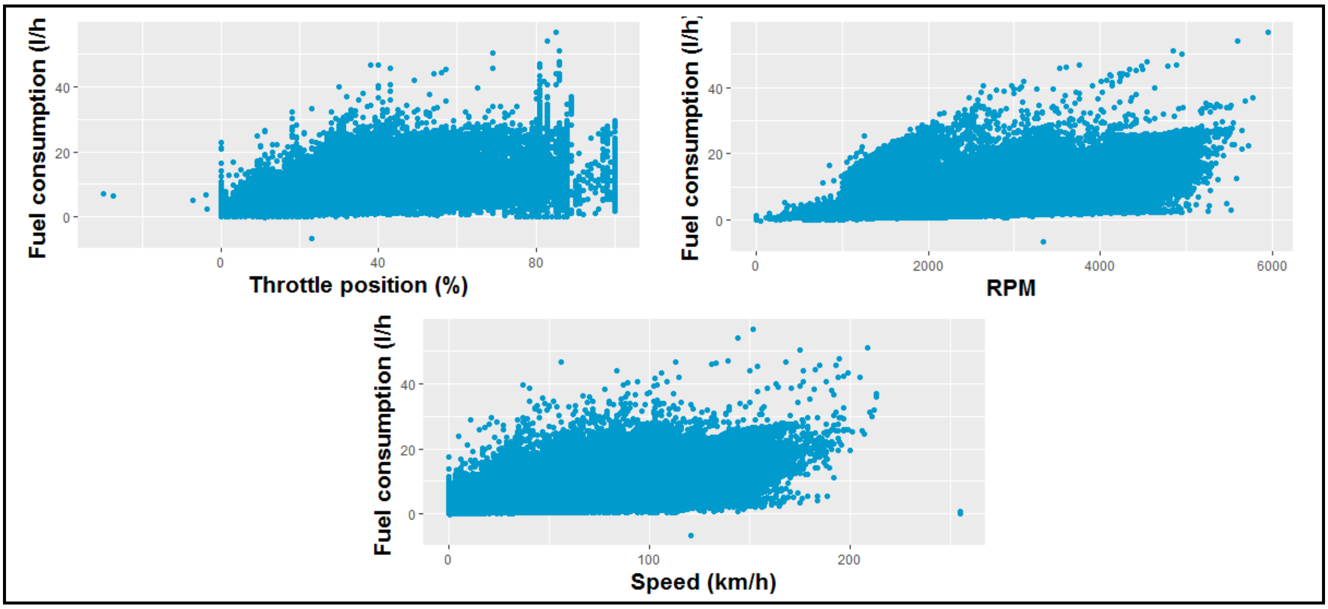


Fig. 33. Scatterplots showing the impact of throttle position, RPM and car speed on fuel consumption.

As anticipated, the focus was on sensors that illustrate the driving style, impact fuel economy, are easily accessible and understandable, and able to provide direct feedback to the driver. To this end, after the reported literature review, I analysed the correlation of FC (computed by enviroCar) with the thirteen OBD-II/enviroCar provided inputs, and finally chose TPS, RPM and car speed. Fig. 33 shows the FC plots against TPS (top right), RPM (top left), and car speed (bottom), for all the analysed available data. As can be seen from Fig. 19 the PPMC between FC and the selected variables is quite high (0.8, 0.82 and 0.75, for TPS, RPM, and speed respectively). Refer to Table 2 for the guidelines to interpret the explicit values of PPMC.

4.1.2 Modelling Fuel Consumption with FL

The Mamdani FL model [176] has been followed in this work. It considers linguistic variables (e.g. very cold, slightly hot) in both the antecedent and consequent parts of the rules. So, in multi-input and a single-output (MISO) systems, the fuzzy IF-THEN rules are of the following form (3):

$$\text{IF } X_1 \text{ is } A_1 \text{ and } \ldots \text{ and } X_n \text{ is } A_n \text{ THEN } Y \text{ is } B, \tag{3}$$

where X_i is input linguistic variable and Y is the output linguistic variable. A_i and B are linguistic values. The standard workflow for the Mamdani model is displayed in Fig. 34, with three main steps:

1. Fuzzification: convert classical (crisp) data into fuzzy data or membership functions (MFs).
2. Inference: combine the fuzzy set definitions of MFs with the applied fuzzy control rules, which can be considered as the knowledge of an expert in the field of application, or base on data analysis to derive the fuzzy output.
3. Defuzzification of the output distribution: convert the linguistic fuzzy output (resulted from the previous step) back to a classical output (real value).

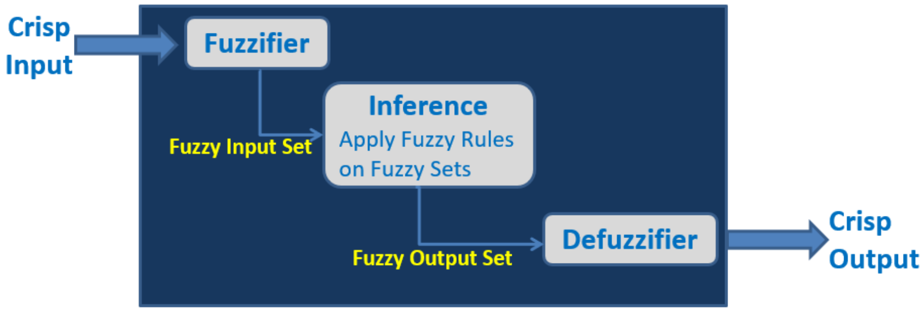


Fig. 34. Fuzzy Inference System of Mamdani model.

In this work, the fuzzy rules were built based on the fuzzy 'AND' operator and all the rule weights are set equal to 1. Defuzzification is done based on the centroid technique. The first step of the process consists of specifying the variable ranges. Observing the available data, the domain for the inputs and the outcome was considered as following, 0-100 (%) for TPS; 0-6000 for RPM; 0-200 (km/h) for car speed; and 0-30 (l/h) for FC.

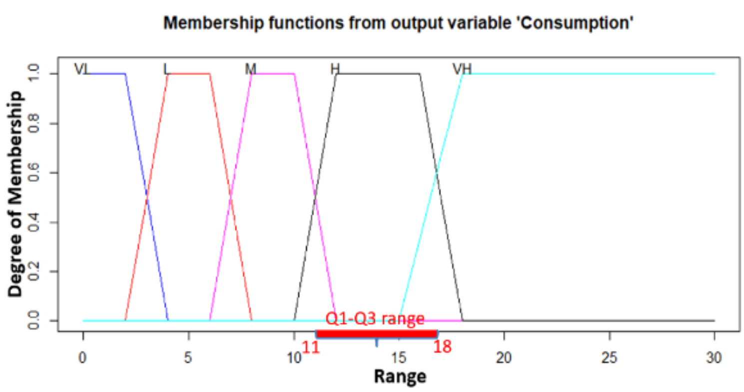


Fig. 35. Example of rule selection.

In order to identify the best solutions, several alternatives have been tested in the definition of MFs of the considered variables. For TPS, after several experimental attempts (with several levels), the MFs determined through expert driver knowledge in [132], have been employed. In terms of the remaining variables (RPM, car speed and FC), their MFs have been defined based on logic, and then they have been refined through trial and error as shown later in this section. This methodology was employed for the definition of the fuzzy rules too.

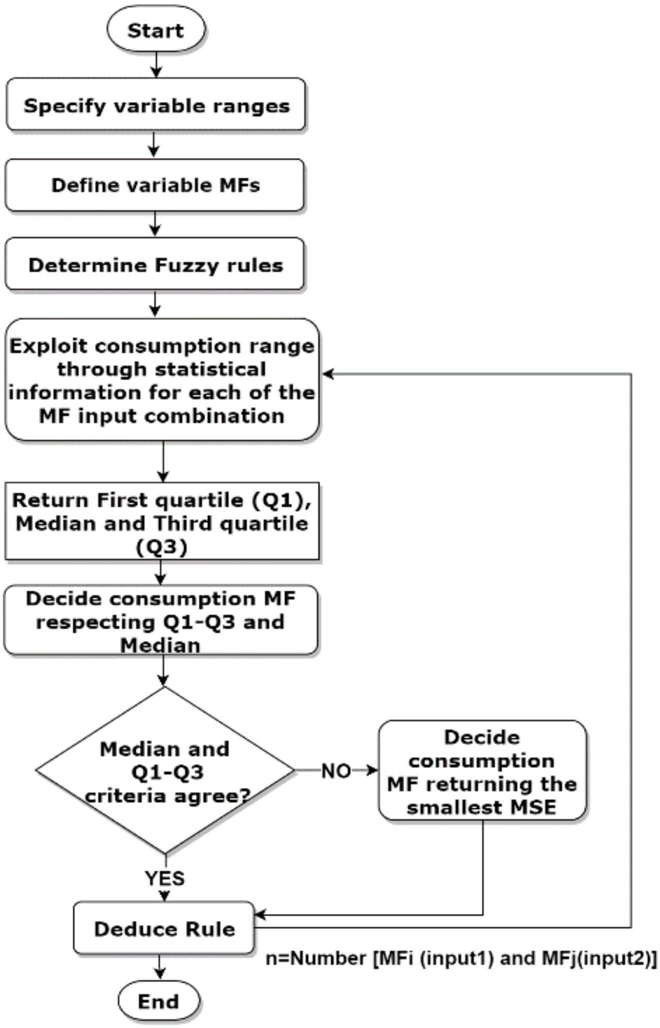


Fig. 36. Followed fuel consumption algorithm.

To define the rules that decide the output MF for each combination of the inputs, the following statistical information have been exploited: First quartile (Q1), which is the middle number between the smallest number and the median of the data set; Median, that marks the middle of the data in the sense that half of the data is less than the median; and the third quartile (Q3), that is the middle value between the median and the highest value of the data set.

This section shows an example of building a rule for FIS1 with TPS and RPM as inputs (described in the next section (A) and its MF plots are illustrated in Fig. 37. The fuzzy values of TPS are defined to be low (L), medium (M) and high (H), as specified through expert knowledge in [132]. Five MFs are used for RPM: very low (VL), low (L), medium (M), high (H) and very high (VH). The MFs of the output are determined to be very low (VL), low (L), medium (M), high (H) and very high (VH). To estimate the FC level when (RPM is H) AND (TPS is H), the FC data is filtered while satisfying the condition: (2500<RPM<4000) AND (TPS>60), and get the following statistical information: Q1= 11.76, Q3= 18 and Median= 14.79. Considering the defined FC MFs ({very low (VL), low (L), medium (M), high (H), very high (VH)}) (Table 5 and Fig. 37 (3)), the Q1-Q3 range overlaps with three FC MFs: M, H and VH), with most of the data is in the H level (Fig. 35). The median is in H as well. Therefore, the subsequent rule has been deduced: IF (RPM is H) AND (TPS is H) THEN FC is H. If the median and Q1-Q3 criteria do not agree (e.g., the median might be in the intersection of two different FC levels), some more processing would be needed. These uncertain cases are handled by measuring the performance of the fuzzy estimator in the two possible output levels for deciding the level that returns the smaller MSE with respect to the reference value. The overall process is illustrated in Fig. 36.

In the following, I present the best performing implementations of FL-based modules for the four possible combinations of the selected input sensors, namely FIS1 with TPS and RPM; FIS2 with car speed and RPM; FIS3 with TPS and car speed; and FIS4 with all the three inputs.

A. FIS1: TPS and RPM

FIS1 involves two inputs: TPS and RPM. The MFs of the two features with their ranges are shown in Table 3 and Table 4 respectively.

Table 3. Membership Functions of TPS (%) input.

Membership Function	**L**	**M**	**H**
Throttle Position (%)	0-40	20-80	60-100

Table 4. Membership Functions of RPM input.

Membership Function	**VL**	**L**	**M**	**H**	**VH**
RPM	0-1000	500-2000	1500-3000	2500-4000	3500-6000

The fuzzification step of the two inputs is made by their MFs. Their MF plots are shown in Fig. 37 (1) and Fig. 37 (2). The fuzzy controller output is derived from the fuzzification of both inputs and outputs, defining the associated MFs. The MFs of the FC outcome, are shown in Table 5, and the corresponding MF plots are presented in Fig. 37 (3).

Table 5. Membership Functions of fuel consumption (l/h) output.

Membership Function	**VL**	**L**	**M**	**H**	**VH**
Fuel consumption (l/h)	0-4	2-8	6-12	10-18	15-30

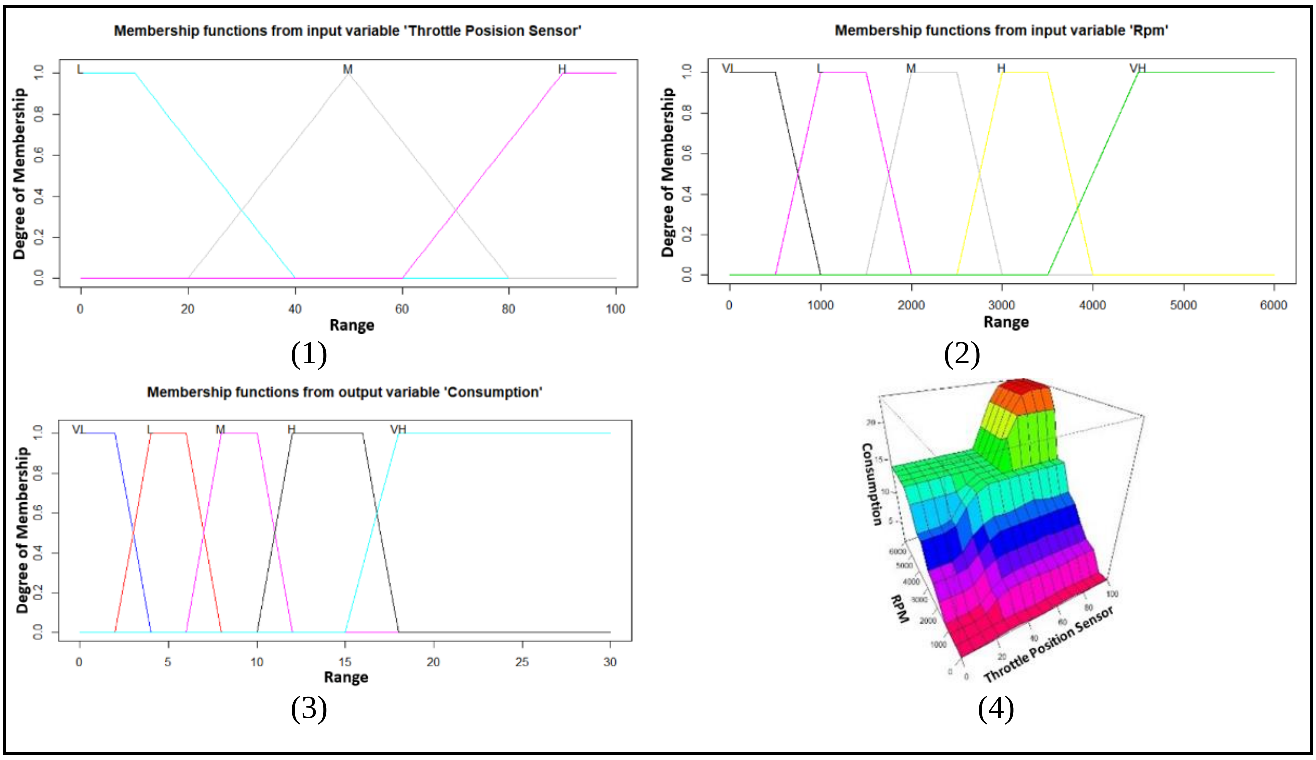


Fig. 37. Membership functions of FIS1 variables (1): TPS, (2): RPM, (3): FC, (4): Whole process mapping by output surface plot for TPS, RPM, and FC.

The fuzzy rules with the 'AND' operator are provided in Table 6. As an example, when TPS is 'M' and RPM is 'VH' then FC is 'H'. The FIS1 three-dimensional view on the data – output surface plot representing the dependency of FC output on the two inputs – is shown in Fig. 37 (4). It is a three-dimensional surface that represents the mapping from the inputs (TPS and RPM) to the output (FC).

Table 6. Fuzzy Rules matrix for FIS1.

RPM / TPS	**VL**	**L**	**M**	**H**	**VH**
L	VL	VL	L	M	H
M	VL	L	M	H	H
H	VL	L	M	H	VH

B. FIS2: Car speed and RPM

The second combination involves car speed and RPM sensors. The MFs of car speed with its ranges are shown in Table 7. Corresponding its fuzzy values, they are defined to be low (L), medium (M), high (H) and very high (VH). The plots of MFs for the inputs and FC output are shown in Fig. 38 respectively: (1) Car speed MFs plot; (2) RPM MFs plot; (3) FC MFs plot. Fig. 38 (4) depicts the generated control surface by the fuzzy system; FC output is plotted against the two input variables (car speed and RPM). The deduced Fuzzy rules are provided in Table 8. For instance, if car speed is ‘H’ and RPM is ‘H’, FC is ‘M’.

Table 7. Membership Functions of car speed (km/h) input.

Membership Function	**L**	**M**	**H**	**VH**
Car speed (km/h)	0-60	40-100	80-140	110-200

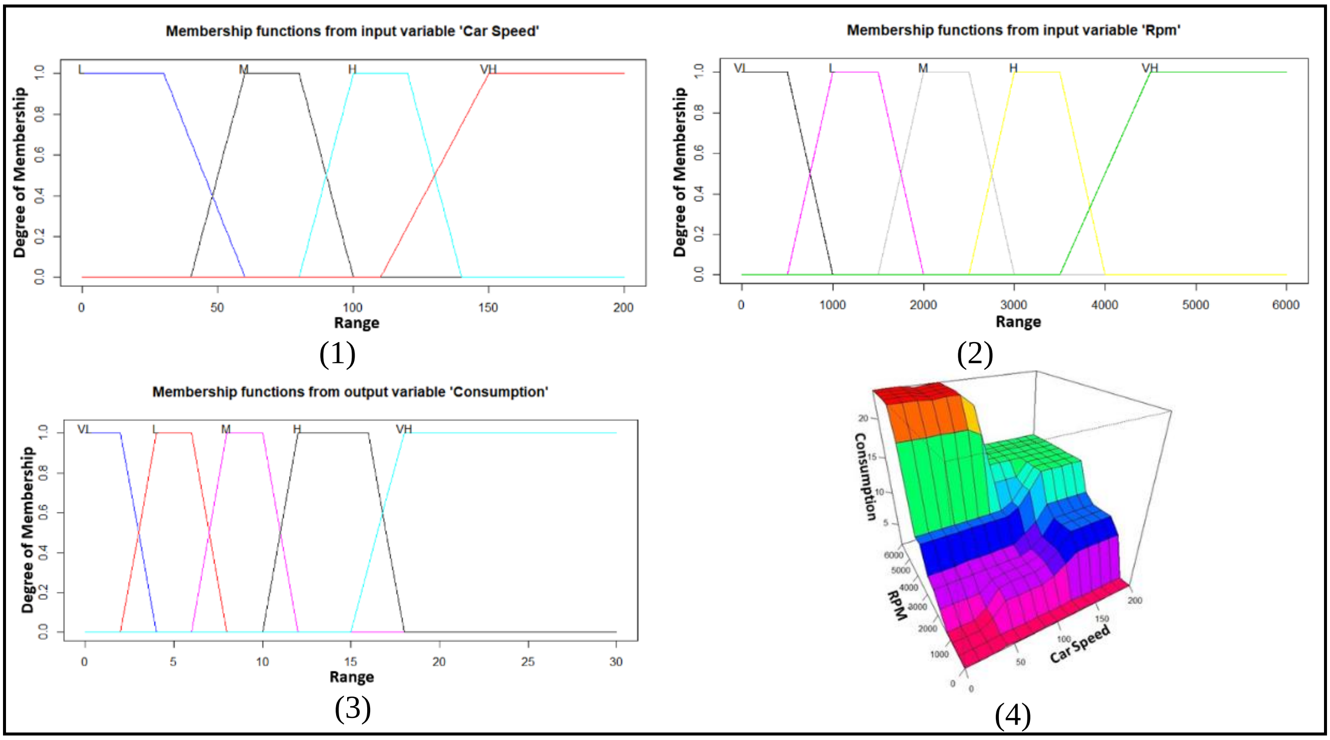


Fig. 38. Membership functions of FIS2 variables (1): Car speed, (2): RPM, (3): FC, (4): Whole process mapping by output surface plot for speed, RPM, and FC.

Table 8. Fuzzy Rules matrix for FIS2.

RPM / **Speed**	**VL**	**L**	**M**	**H**	**VH**
L	VL	VL	L	M	VH
M	VL	L	L	M	VH
H	VL	L	L	M	H
VH	VL	M	M	H	H

C. FIS3: Car speed and TPS

The combination of TPS and car speed is modelled in FIS3 (Fig. 39). The 3D interpretation of the FIS3 input-output relations is illustrated with an output surface in Fig. 39 (4). The fuzzy rules – learned from data for the combination of both inputs with the 'AND' operator, are provided in Table 9. I do not report the MFs, as they are the same as in the previous models; TPS in Fig. 37 (1), RPM in Fig. 37 (2), car speed in Fig. 38 (1) and FC in Fig. 37 (3).

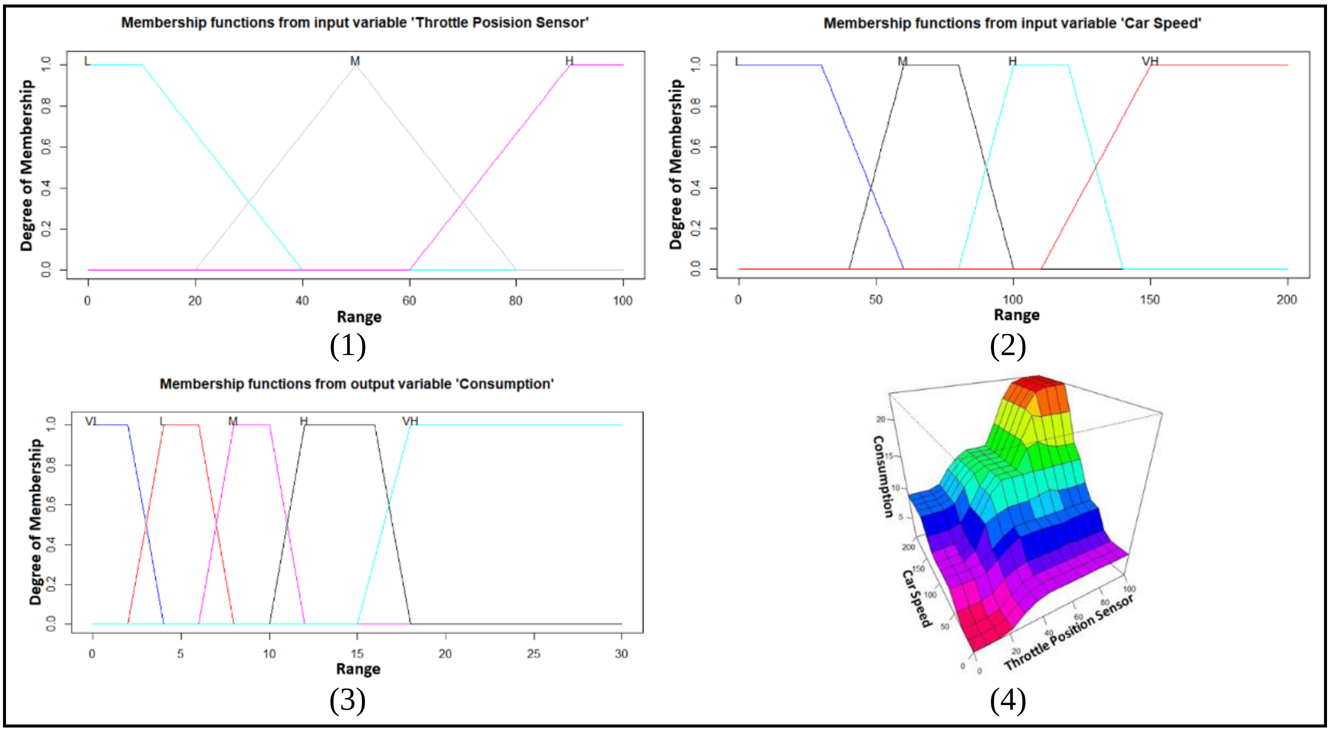


Fig. 39. Membership function for the inputs and output of FIS3, (1): TPS, (2): Car speed, (3): FC, (4): Whole process mapping by output surface plot for TPS, speed, and FC.

Table 9. Fuzzy Rules matrix for FIS3.

TPS \ Speed	**L**	**M**	**H**	**VH**
L	VL	L	L	M
M	L	M	H	H
H	L	M	H	VH

D. FIS4: Car speed, RPM and TPS

The last FIS involves all the three inputs, for which the MFs were shown previously. The sixty possible combinations of the MFs with the 'AND' operator were studied, with TPS ({L, M, H}), RPM ({VL, L, M, H, VH}), and car speed ({L, M, H, VH}) respectively. After the synthesis step – deciding, with the suggested approach (Fig. 36), the following seventeen rules have been deduced:

1. if RPM is VL then FC is VL
2. if RPM is L and TPS is L and Speed is (L or M or H) then FC is VL

3. if RPM is L and TPS is L and Speed is VH then FC is L
4. if RPM is L and TPS is M then FC is M
5. if RPM is L and TPS is H and Speed is (L or M) then FC is L
6. if RPM is L and TPS is H and Speed is H then FC is VH
7. if RPM is L and TPS is H and Speed is VH then FC is H
8. if RPM is M and TPS is L then FC is L
9. if RPM is M and TPS is M then FC is M
10. if RPM is M and TPS is H then FC is M
11. if RPM is H and TPS is L then FC is L
12. if RPM is H and TPS is M and Speed is L then FC is M
13. if RPM is H and TPS is M and Speed is (M or H or VH) then FC is H
14. if RPM is H and TPS is H then FC is H
15. if RPM is VH and TPS is L then FC is M
16. if RPM is VH and TPS is M then FC is H
17. if RPM is VH and TPS is H then FC is VH

The reduction to seventeen rules was obtained by synthesizing all the sixty resulting rules. For instance, when RPM is 'VL', FC is always 'VL', independent of the level of TPS and speed. The 12 rules where RPM is 'VL' could thus be summarised by the first rule above (if RPM is VL then FC is VL). In general, the rules highlight a higher impact on FC from RPM, compared to speed and TPS. Also, I argue that TPS is the second informative ones for the model. As an example, in rules (15) -(17), in case having 'Very High' level for RPM, the output is based on the TPS value: 'Low' TPS implies 'Medium' FC; 'Medium' TPS implies 'High' FC; 'High' TPS implies 'Very High' FC.

4.1.3 Experiment Results and Discussion

The Fuzzy models were developed using the "FuzzyToolkitUoN" R package [177]. The experimental tests were conducted at the same computational environment, on an 8 GB RAM and i7-770 CPU desktop PC. For each one of the four FL models, every input sample was processed within 0.01 seconds (estimator's latency time experienced by a user). FC output values of the fuzzy inference systems (FISs) were then compared with the consumption reported in the enviroCar database.

Table 10. Model performance comparison.

Track \ Model	**FIS1**(TPS & RPM)		**FIS2(**Speed & RPM)		**FIS3**(TPS& Speed)		**FIS4** (3 inputs)	
	R^2	**MSE**	**R^2**	**MSE**	**R^2**	**MSE**	**R^2**	**MSE**
All gasoline data	0.62	5.8	0.47	8.25	0.64	5.52	0.71	4.69
Same car on different paths	0.85	1.75	0.53	4.75	0.76	3.53	0.86	1.52

Table 10 provides the differences between the developed models' outcome and the target FC provided by the enviroCar database in terms of R squared or squared correlation coefficient (R^2) and the MSE. The model with three inputs (FIS4) is the best one, with the highest R^2 (0.71) and the lowest MSE (4.69). FIS1 and FIS3 return good results, with R^2 equal to 0.62 and 0.64, respectively, and with MSE equal to 5.8 and 5.52, respectively. FIS2 is not as accurate as of the other models, achieving the lowest R^2 (0.47) and the highest MSE (8.26). The absence of TPS in FIS2 suggests its importance.

In order to check the effect of some of the other factors illustrated in Fig. 1, and not captured by the involved sensors, the models were evaluated on data from the same car. 111 different driving tracks were considered with 47,076 measurements (after ignoring incomplete records) for a gasoline car with the following characteristics: Volkswagen manufacturer, 9N model constructed in 2009. In this case, the results show higher accuracy than in the general case (when working on different car types), for all the four models, with higher R2 values 0.85, 0.53, 0.76 and 0.86 and lower MSE values 1.75, 4.75, 3.53, and 1.52 for FIS1, FIS2, FIS3 and FI4 respectively (Table 10). This result suggests the importance of considering also other factors (e.g., car characteristics such as engine size, number of cylinders, engine displacement) in estimating FC in case a higher accuracy is needed.

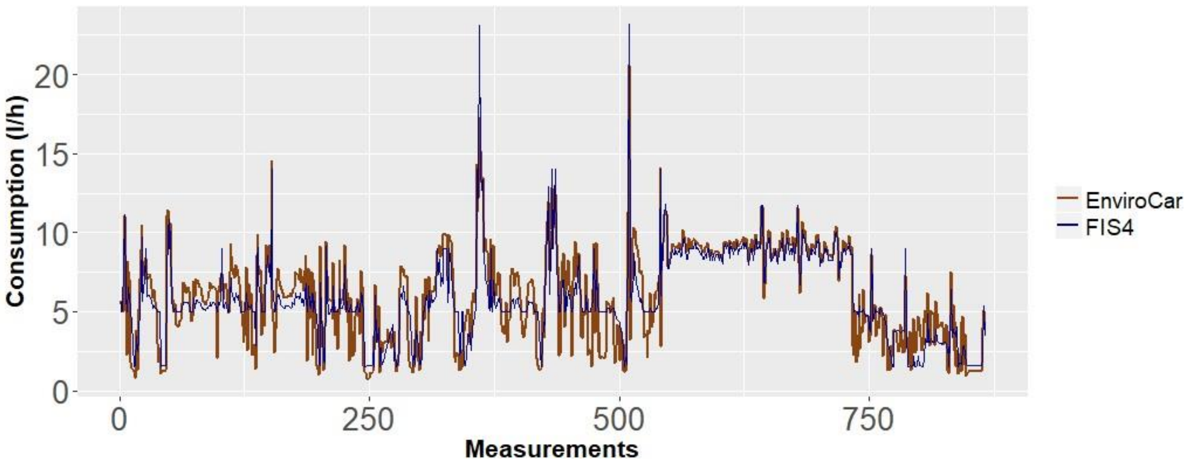


Fig. 40. Evaluation of FIS4 with a track.

Fig. 40 shows the time evolution of the obtained estimation with the FIS4 model versus the actual enviroCar data in a sample track. The predictor is able to follow the main peaks of the signal (even without the possibility of having a direct input about acceleration/deceleration), while it has difficulty in capturing small variations around the average value around 5 l/h. This work argues that the proposed FL module might be improved in those cases by introducing other inputs, for instance, the OBD-II calculated engine load.

As anticipated, one of the expected benefits of employing FL consists in the possibility of providing coaching feedback to the driver (the link between outcome with inputs) and by the fact that the rules are applied to input values that – at least in this work case – can be directly controlled by the user. As a simple example, upon the 14th rule of FIS4, if a “High” FC is obtained because of a “High” RPM and a “High” TPS value, the player could be advised to decelerate or to upshift the gear.

4.1.4 Methodology Recap

In this chapter, I have focused on FL as a baseline approach in order to estimate automotive driver performance in terms of FC, based on TPS, RPM and car speed that are read from the OBD-II interface. FL is a promising technique for driving style analysis [85], that gives the possibility of distinguishing between different performance factors, so to give coaching feedback to the driver.

Four fuzzy inference systems with all the possible combinations of the inputs have been modelled: FIS1 (TPS and RPM), FIS2 (speed and RPM), FIS3 (TPS and speed) and FIS4 (with all the three inputs). They have been defined on the basis of naturalistic driving data (733,274 measurements) that was not calibrated for a specific car model, which made the work challenging. Data were taken from the enviroCar database (described in chapter 3), and the FC predictions were compared with theirs, via regression analysis for a track.

The MFs and the fuzzy rules for the models have been defined based on literature expertise, data analysis and trial and error. Results indicate a R^2 best value of 0.71 and 4.69 for MSE, with FIS4 in the regression evaluation process with the observed consumption provided by enviroCar. The deduced fuzzy rules show that RPM is the strongest FC predictor, followed by TPS and car speed.

FIS1 and FIS4 models achieve good results when dealing with only one vehicle, confirming the importance of considering other variables beyond those I got from the OBD-II port, such as vehicle characteristics (e.g., engine size and displacement). Finally, the proposed approach achieves a RT performance (0.01 s. response time), which is needed for gaming. These results are promising, and further work is required to improve estimation accuracy.

I also argued that the proposed FL framework can include the indirect uncertainties in measurement and it can make possible to lower the particulate matters (PM) and FC simultaneously using MAF and in-cylinder pressure. The fact that in addition to FC, the PM has a more important role in the vehicle overall emission comparing to harmful gases NO_x, HC and CO, which by using the three-way catalyst and stoichiometric combustion can be comfortably reduced down to the European standards. Measuring the PM concentration levels is directly not feasible. Employing novel Gasoline Direct Injection (GDI) technology in modern engines for this purpose still, face a number of challenges in acquiring accurate PM data. The concentration level, indicated by specific particulate matter by mass (ISPMM) in g/kWh (see eq. (4)), is directly related to the measurement of fuel mass flow (MFF) rate and MAF rate in g/min [178][179].

$$SPMM = \frac{CAM \ \cdot 10^{-6} \cdot 10^{3} \cdot (MAF + MFF) \cdot 60}{P_i \cdot \frac{MMEx}{22.7L} /mol} \tag{4}$$

To lower the PM concentration, it is necessary to normalize the FC and PM emissions values by engine power outputs (P_i), which is, in turn directly related to the in-cylinder pressure measured via ECU in every engine cycle, e.g. 10ms. The proposed FL approach can include such indirect measurement uncertainties for the in-cylinder pressure, MAF, and MFF due to other environmental variables and inherent measurement inaccuracies in current sensor technology. Therefore, it can give more accurate results in reducing both the FC and PM in case of measurement uncertainties, compared to other AI or ML approaches, which heavily depends on direct sensory data and cannot cope well with measurement uncertainties.

4.2 Methodology 2: Random Forest versus Fuzzy Logic of Methodology 1 in Fuel Consumption Estimation

In this chapter, I am interested in exploring if better quantitative FC estimations than the achieved ones via FL model (4.1) could be obtained with a ML technique, selecting random forest (RF) which is introduced in 2.4.2. Hence here, I am analyzing two types of algorithms for driver performance modelling: FL described in 4.1 and RF. The former is a method of reasoning that resembles the human reasoning used in the development of human-like capabilities for AI. It is a promising technology for driving style analysis [85]. It has the ability to deal with incomplete information [28][126], to distinguish between different performance factors by matching any set of input-output data, transferring human knowledge and expertise into a mathematical model by means of if-then rules [127], so to give coaching feedback to the driver about his or her performance (see 4.1). The latter is one of the ensemble learning methods, a powerful ML algorithm, that is quickly trainable, performs implicit feature selection by providing an indicator of feature importance.

The remainder of this chapter is organized as follows: Section 2.4.2 introduces RF; Section 4.2.1 reports on the work in the modelling stage of an IoT-enabled RESG using RF technique; Section 4.2.2 presents the experiments to compares and assess both proposed models with FL and RF approaches; 4.2.3 recaps this methodology.

4.2.1 Modelling Fuel Consumption with Randon Forest

For implementing the RF model, the '*RandomForestRegressor*' of the '*sklearn.ensemble*' python library [180] has been used. The most important settings are the number of trees (*n_estimators*) in the forest and the number of features considered for splitting at each leaf node (*max_features*). For the developed RF, the following six hyperparameters have been adjusted,

(1) *n_estimators*: number of trees to build in the forest.
(2) *max_features*: max number of the considered features for splitting a node.
(3) *max_depth*: max number of levels in each decision tree.
(4) *min_samples_split*: min number of data points placed in a node before the node is split.
(5) *min_samples_leaf*: min number of data points allowed in a leaf node.
(6) *bootstrap*: a method for sampling data points (with or without replacement).

Random search and grid search have been used to determine the optimal parameters of the RF model. They have been implemented using '*RandomizedSearchCV*' [181] and '*GridSearchCV*' [182] respectively. Both methods are optimized via 10-fold cross-validation (CV) for reducing the chance of overfitting. First, I followed a random sampling approach during fitting, using the random search technique. I tried 60 different model settings (fitting 10-Fold CV for each of 6 candidates). At each iteration, the approach uses a random combination of the features to sample a wide range of values to end up with the best model. This helps in narrowing down the range for each hyperparameter values, that have been define in a grid of their ranges as following n_estimators [400, 500, 800, 1000], max_features [auto, sqrt], max_depth [10, 20, 30, 40, 50, 60, 70, 80, 90, 100, None], min_samples_split [2, 5, 10, 14], min_samples_leaf [1, 2, 3, 4, 8] and bootstrap [True, False]. The best model found by this approach is RF_1 [n_estimators = 800, max_features = sqrt, max_depth = 90, min_samples_split = 5, min_samples_leaf = 4, bootstrap = False].

In a trial to further improve the results, a grid search approach was used then to evaluate all combinations based upon the most promising ranges that are found in the random search: n_estimators [700, 800, 900], max_features : sqrt, max_depth [80, 90, 100], min_samples_split [4, 5, 6], min_samples_leaf [3, 4, 5] and bootstrap : True. The grid search fitting suggests the optimal model is RF_2 [n_estimators = 800, max_features = sqrt, max_depth = 100, min_samples_split = 6, min_samples_leaf = 3, bootstrap = False]. This RF model has been followed then.

4.2.2 Experiment Results and Discussion

The available dataset was divided into a training set (80% of the data) and a testing set (20%). Table 11 shows the accuracy of out-of-sample performance, the likely performance of both models FL and RF (derived from the grid search). MSE and R^2 statistical metrics have been used to assess the prediction performance. Comparing the two studied techniques, RF performs much better than FL, both in terms of R^2 (0.896 vs. 0.65) and MSE (1.506 vs. 4.74). Fig. 41 shows the fits of both models FL (a) and RF (b) for the testing set (171,484 samples). Fig. 42 visualizes the time evolution of FL (a) and RF (b) versus the actual enviroCar FC data in a sample track. Considering the computation performance, which is crucial for gaming, I highlight that both FL and RF-based models can achieve a RT execution, with latencies within a millisecond.

Table 11. Comparison of performance for the FL and RF.

Model	**FL**	**RF**
MSE	4.745	1.506
R^2	0.650	0.896

Observing the fuzzy rules to interpret the model described in the previous chapter Section D, I assumed that RPM is the most important FC predictor, followed by TPS and car speed [4]. For a more accurate picture, this study takes advantage of the RF feature importance interpretation tool, by measuring the prediction strength of each variable. For computing the feature importance in RF, two strategies were applied. First, I relied on the most common mechanism – the 'mean impurity decrease' (MID), default variable importance in '*scikit-learn*' [136]. Then the '*permutation importance*' or '*mean decrease in accuracy*' (MDA) has applied [183]. The permutation importance technique is broadly applicable to its reliability [184].

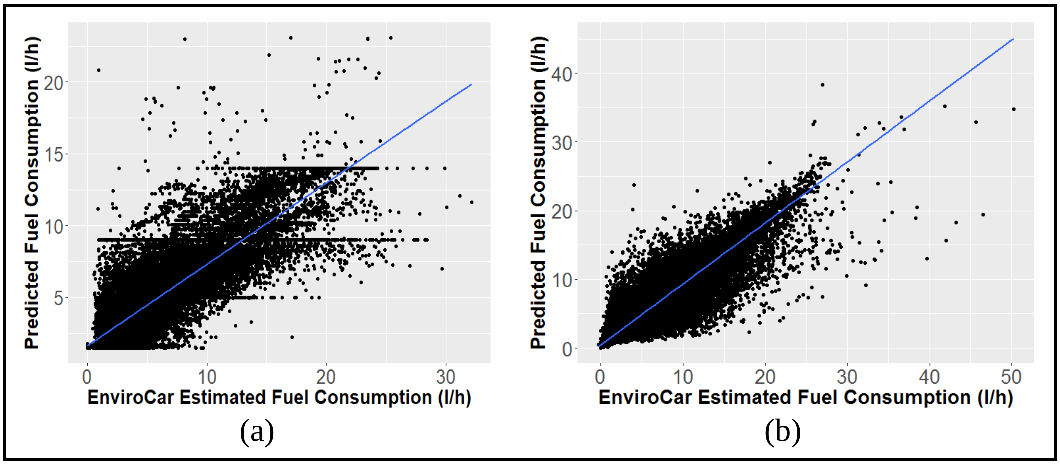


Fig. 41. Comparison of the model fits (a) fuzzy logic and (b) random forest.

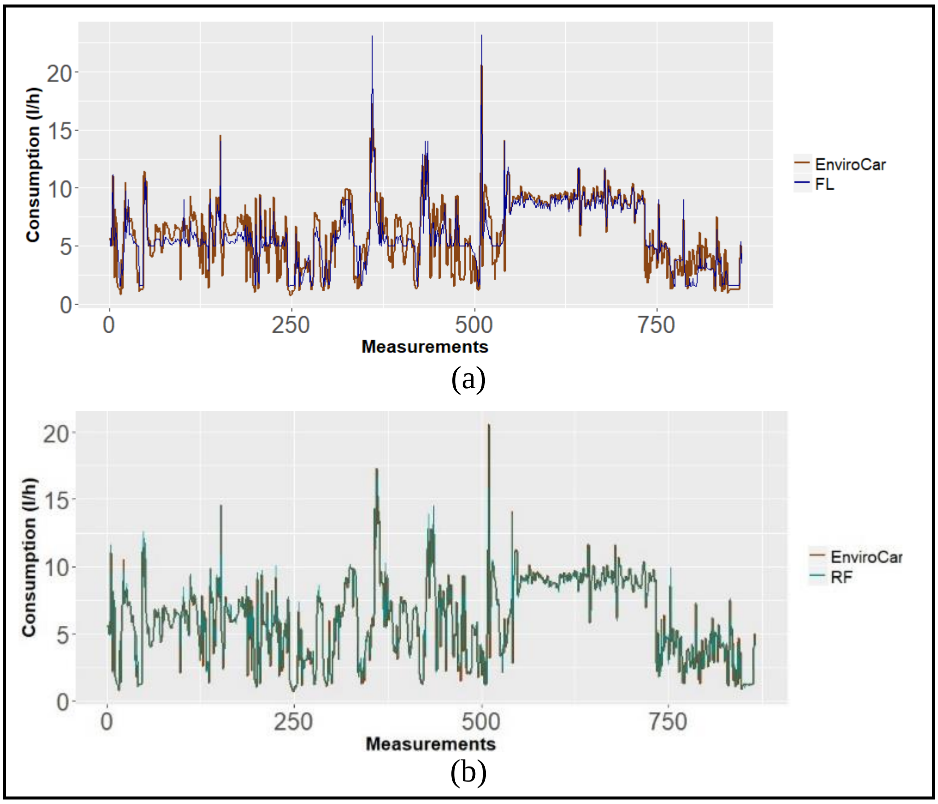


Fig. 42. Track data evaluation with the models FL (a) and RF (b).

According to MID, car speed is the most influential feature of the variables in the FC model, followed by TPS and RPM respectively (Fig. 43 (a)). Comparing this with the results of the FL model in 2.4.1 Section D, it has been observed that RPM is substituted by car speed as the strongest FC predictor.

According to the MDA (Fig. 43 (b)), RPM is again the most important FC predictor, followed by TPS and car speed, confirming the main findings of the suggested FL-based previous work at the end of section 4.1.2.

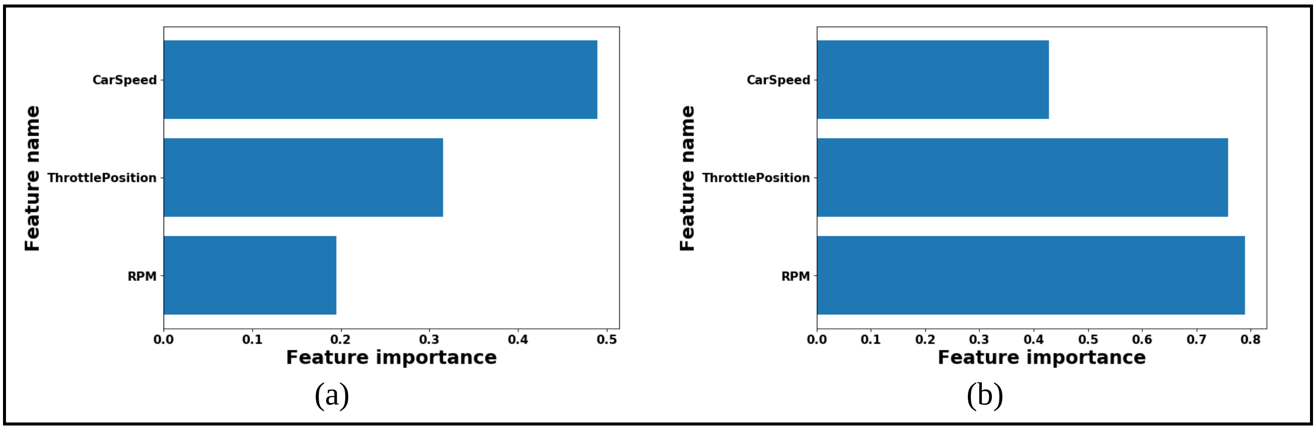


Fig. 43. Variable importance plots for the RF regressor with (a) MID and (b) MDA.

This analysis is key to a SG designer for proper prioritisation of the predictors while supplying the drivers with instantaneous feedback to improve their efficient behaviour. Taking the previous example, based upon the 14th rule in D, if a “High” FC is obtained because of a “High” RPM and a “High” TPS value, then the system should give priority to the gear shift message, and then to the deceleration suggestion.

4.2.3 Methodology Recap

This chapter reports my experience when exploring two types of algorithms for FC modelling for use in IoT-enabled driving SGs: FL and RF. I spent a significant amount of time for the data analysis needed, to build the MFs and the fuzzy rules for the FL model. Note also, the application of hyperparameter tuning techniques (random search and grid search) along with the 10-folds CV for optimal RF model configurations was computationally expensive.

A FL model provides linguistic understandable feedback, concerning input to output links. It enables specifying the likely performance causes, based upon the deduced fuzzy rules, as presented in D, resembling human reasoning, which is useful to support a player’s performance improvement, in a SG. RF, on the other hand, is demonstrated to be able to predict FC much more accurately, with higher R^2 (0.896 vs. 0.65) and lower error (1.506 vs. 4.745) (Table 11), better capturing the trends in data. Moreover, RF provides a quantitative indicator of feature importance, according to which RPM is estimated to be the most informative signal, followed by throttle position and car speed.

These results suggest an opportunity to take advantage of both models, also considering that they both can support RT execution, with latencies within a millisecond. I thus argue that the combination of the two models can provide valuable information usable in reality-enhanced SG. FC estimations with good accuracy are needed as continuous sequential determinations of FC can be realised by a game design directly as an ongoing updated score, or as the energy of the player, or to activate bonuses or maluses, or to facilitate reaching a higher gaming level, etc. (e.g., [22][32]). FL rules provide very good hints on the causes of the driver's performance level, but RF allows achieving a much better quantitative estimation. Moreover, the RF feature importance tool gives important insights into the prioritization of the predictors, which in turn enhances the communication of the outcomes of the FL rules.

4.3 Methodology 3: Module 1 Towards a Reality-Enhanced Serious Game to Promote Eco-driving in the wild

In this chapter, I attempt to improve the estimations of FC with RF that it is presented in the previous chapter (4.2). During a drive, the values of car speed, TPS and RPM are captured periodically on a smartphone connected via Bluetooth to the vehicle OBD-II interface through an adapter (Fig. 2).

Given the effectiveness of monitoring a driver's behaviour [34], the module supplies the drivers with direct feedback considering what actions they could do toward eco-friendlier attitude via (1) TPS, (2) car speed, (3) RPM and (4) the estimated car jerks (changes in acceleration with respect to time). Significant changes in these act as events to be detected that affect FC, e.g., detecting drivers overtaking. The instant feedback is categorised based upon the literature.

On the other hand, it's beneficial to use gamification tools to score efficient behaviour of drivers after each trip such as on a scale of 100 (100 is the best possible score), and "gamify" it – through proper game mechanics – to rank the performance of a driver against drivers' peers. The more the driving becomes efficient, the higher the score (closer to 100) the driver earns. While the lower the score, the more the driving is inefficient and aggressive. Based on the final score, the driving pattern can be classified (such as fuel 'Saver') besides, a summary report for the driving pattern which could be useful.

In eco-driving profiling, fuel-efficiency is the most important metric. Although, other factors can also lead to an increased FC in addition to driving styles (check 1.2.3), such as traffic congestion, atmospheric pressure and weather conditions, or a higher engine load occurs because of the car configuration (e.g., use of heated seats, headlights) [4], it's difficult to acquire and fuse the sensor data for these.

The remainder of this chapter is organized as follows: 4.3.1 describes the methodology: RF model for prediction the consumption of fuel and the event detectors needed for the recommendation while driving, besides the eco-driving classification; 4.3.2 presents a case study and discussion; 4.3.3 summarises this chapter.

4.3.1 Methodology

Fig. 44 depicts the followed methodology. The inputs are periodically captured from the vehicle OBD-II interface, through an OBD-II adapter and delivered to a driver's smartphone via a Bluetooth connection. For the analysis, 8726 different gasoline tracks were considered, with 983291 measurements for gasoline engines that were recorded mostly in Germany in the period 2012-01-01 – 2016-06-15. In order to retrace the journey of each car, an ID number identifies each of the direct features, besides the timestamp extracted from enviroCar smartphone application.

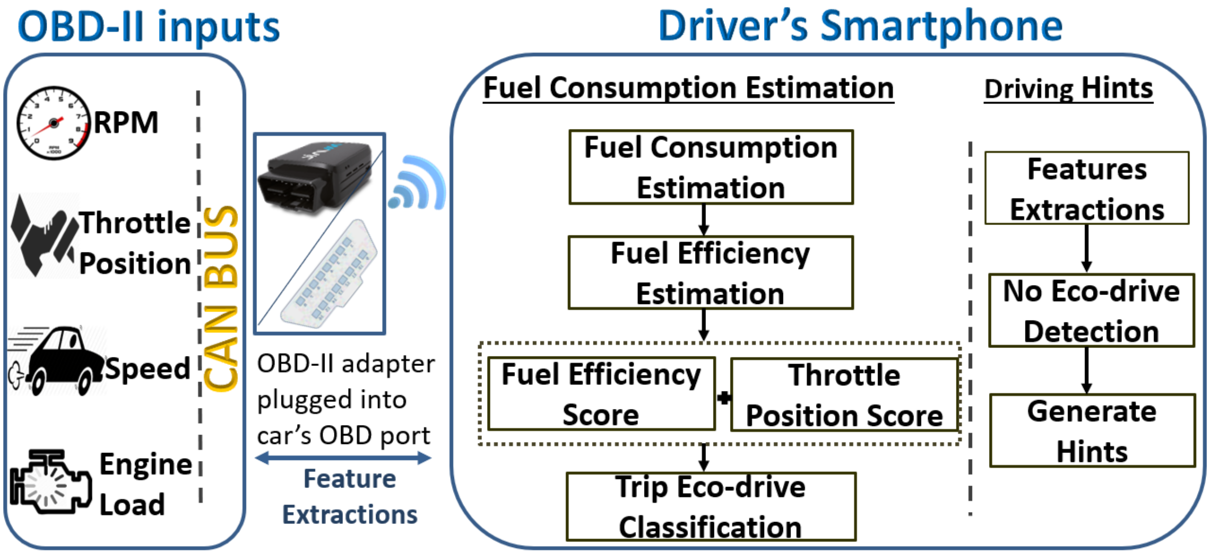


Fig. 44. In-car features extraction via OBD-II system for both processes (running in drivers' smartphones): (ii) FC estimation and (ii) RT driving feedback when inefficient manoeuvres are detected.

A. Improving Instant Quantitative Fuel Consumption Estimation and Trip Categorization

Fig. 45 depicts the followed methodology.

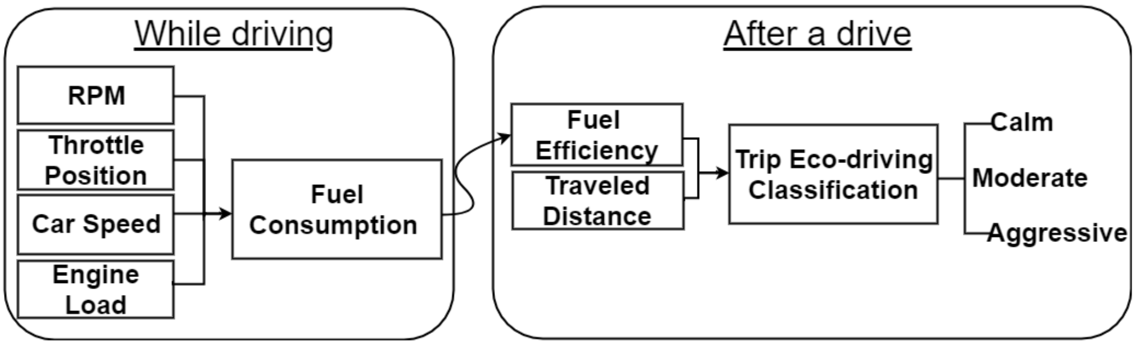


Fig. 45. FC estimation and driving classification methodology.

i. Instant Quantitative Fuel Consumption Estimation

In the previous chapter, I improved the FC predictor modelled by FL in 4.1 with RF technique for gaming via three vehicular signals TPS, RPM and car speed (section 4.2.1 [5]). FC is impacted by other factors in addition to driving styles (see 1.2). Considering other variables (even if not directly controllable by the driver), could help in increasing the model accuracy with much more complex models design. Consequently, I involved the calculated 'engine load' that is sensed from OBD-II, as a further FC predictor. This variable (ranging from 0% to 100%) measures how much air and fuel are sucking into the engine. The more the engine is loaded (close to 100%), the more the fuel is burned (my primary data analysis 3.5.6, Fig. 26). This positive correlation between the engine load and the consumption of fuel is clear with a high PPMC equal to 0.85 in Fig. 19.

Likewise in 4.2.1, for implementing the RF model, I used the ‘*RandomForestRegressor*’ of the ‘*Sklearn.ensemble*’ python library [180], dividing the local dataset in 80% learning and 20% testing. I have adjusted its two most important settings, (1) the number of trees in the forest ‘n_estimators’ and (2) the number of features considered for splitting at each leaf node ‘max_features’. In addition to ‘max_depth’, which is the max number of levels in each decision tree.

ii. Trip Eco-driving Categorization

The more the driving becomes eco-friendly and efficient, the higher the score (closer to 100) the player earns. The lower the score, the more the driving is inefficient and aggressive. I relied on the “Elbow” method as a hint in the decision of the number of driving style groups. Fig. 46 depicts the “Elbow” chart for FC combined with RPM, TPS and car speed respectively. RPM, TPs and car speed affect strongly FC and they are affected by driving style (check 4.1). The elbow – the point of inflection on the curve – for both line-charts (Fig. 46) relevant to RPM-FC and car speed-FC, suggests that two categories are enough to represent driving behaviour while the line-chart of TPS-FC suggests selecting three classes as a first choice. In this approach, I considered three classes as suggested by TPS-FC, classified a trip as fuel ‘Saver’, ‘Typical’ or ‘Careless’.

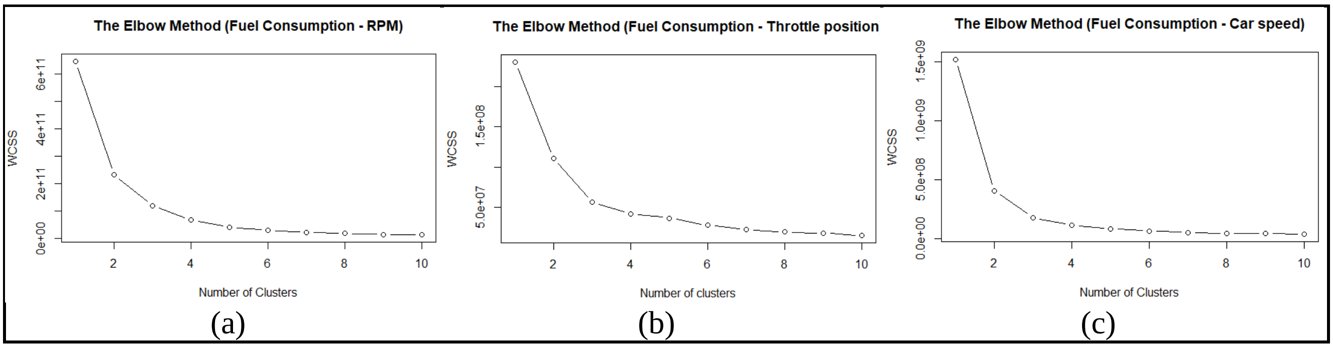


Fig. 46. Elbow method for a figure of a number of driving groups for FC with RPM (a), TPS (b) and speed (c) respectively.

The fuel-efficiency relates distance travelled by a vehicle and the amount of fuel consumed (Eq. 5). Manufacturers give a fuel economy figure for new cars in litres per 100 km under ideal conditions for urban, extra-urban (e.g., higher speeds) and combined (a mixture of the two), which is difficult to achieve.

$$\text{Fuel} - \text{efficiency} = \frac{\text{kilometres travelled}}{\text{total trip fuel consumed}} \tag{5}$$

When a trip starts, the driver gets 0 points, and then points are earned as the trip progresses. For estimating the score of the achieved fuel-efficiency, I followed the proposed approach given by [185] (Fig. 47). The algorithm compares the fuel-efficiency achieved by a driver with the current maximum recorded fuel-efficiency value by any previous driver, who is driving a similar car model. This avoids figuring out the maximum fuel-efficiency for every car model.

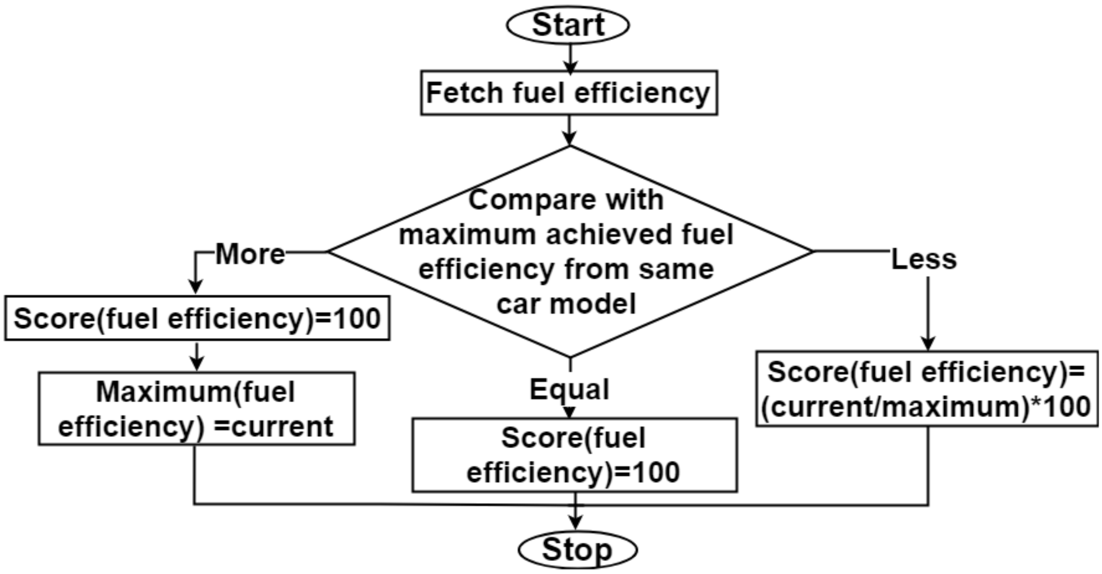


Fig. 47. Gasoline fuel-efficiency score algorithm [185].

B. No Eco-driving Events Detection

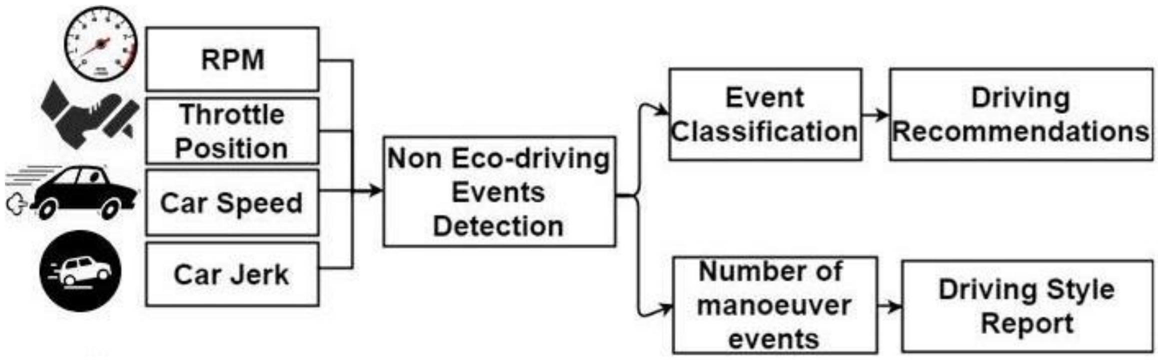


Fig. 48. Instant driving recommendation methodology for inefficient driving patterns.

During a drive when aggressive/inefficient driving manoeuvres are identified via RPM, TPS, car speed and car jerk, the module triggers driving advice, on what actions the driver could do, to better control the fuel economy (Fig. 48). The FC metric was not involved for the robustness of the system, due to the difficulty of its assessment (1.2.3), since it is affected by several internal and external factors such as car characteristics (e.g., engine size, number of cylinders, engine displacement), car weight (e.g., number of passengers), car configuration (e.g., tyre pressure, entertainment equipment use), and driving environment [4].

a) *TPS* (sensed from OBD-II), ranging from 0% to 100%: It is so named because it regulates the air and fuel intake into the engine, making it run slower or faster – the more the accelerator pedal is pressed (TPS value is close to 100%), the more fuel and air will be supplied to the engine and ignited (4.1.1, Fig. 19). It is one of the directly controlled parameters by the driver, reflecting the driver's preferences or driving habits in dealing with the accelerator pedal. The driving style was categorised based upon this event detector into three classes calm 'C', moderate 'M' and aggressive 'A', when its value ranges in 0-39, 40-59 and 60-100 respectively (Table 12). If the driver is pressing the accelerator pedal strongly (class 'A'), the following advice is going to be raised "Release the accelerator pedal gradually, too much fuel is supplied to the engine".

Table 12. Driving classification with the throttle position sensor.

TPS (%)	Class	Recommendation
0-39	C	-
40-59	M	-
60- 100	A	Release the accelerator pedal gradually, too much fuel is supplied to the engine

b) *RPM* (sensed from OBD-II), expressed as the number of revolutions per minute: A strong positive correlation exists between RPM and FC: the higher the RPM, the more the fuel is consumed (3.5.4, 4.1.1) [4][5]. For more economic fuel, it is recommended to shift gear to stay within 2000-3000 RPM, staying less than 3500 RPM [57]. This differs from a car to another, considering several features such as engine type, engine characteristic. Table 13 shows the driving classification with this indicator. For RPM value lower than 2000, the driving was classified as calm 'C'. If the engine revolves between 2000 and 3000 times per minute, the driver is considered moderate 'M'. For engine RPM value greater than 3000, careless 'CA' driving style is considered, then the module advises the player to shift to a lower RPM with "Slow down or shift up a gear to save fuel". Noting that the level of RPM's control depends on the transmission type, where full control over RPM could be achieved with manual transmission via gear selection for speeding up or slowing down the engine (RPM is a factor of speed and gear setting [63]).

Table 13. Driving classification with RPM.

RPM	Class	Recommendation
<2000	C	-
2000-2999	M	-
3000-	CA	Slow down or shift up a gear to save fuel

c) *Car speed* (sensed from OBD-II), measured in km/h: It is a vital input for driving analysis – a strong predictor of crash involvement, and yet it is positively correlated with FC (3.5.5, 4.1.1) [4]. Adding to that, overspeed is a crucial metric to characterize the driver safety compliance toward himself or herself, the passengers concerning the surrounding speed limit. Overspeeding manoeuver events are triggered if the vehicle's speed is greater than the legal speed limit. However, the speed limit depends on the road category (e.g., motorway, congested urban road). In order to supply the player with instantaneous speed limit values through the game, the maximum legal speed was requested through a web service access, based on OpenStreetMap[1] (OSM) with GPS latitude and longitude for each sample as RT context information ("maxspeed" tag of the JSON response). Table 14 presents the driving classification with this indicator. The driver is classified by moderate' if he or she obeys the legal speed limit: car speed is less than or equal to the maximum speed. In case the car's speed reaches the allowed speed limit, the gaming module triggers a notification with the context "The legal speed limit is reached". Yet, when exceeding the speed limit, the game warns the driver via "overspeeding, slow down for safety and fuel-saving" with the exceeding value (the difference between current speed and the speed limit). In this situation, the driver is considered aggressive 'A'. It has been experienced that the legal speed limit values for some samples are not provided by OSM. This might be the reason for low GPS-accuracy of the measurements caused by a poor signal quality of the GPS satellite (e.g., losing GPS reception in a tunnel).

Table 14. Driving classification with car speed (CS: current speed, MS: maximum speed read from OSM).

Speed(km/h)	Class	Recommendation
CS<MS-5%	M	-
MS-5%≤ CS≤ MS	M	The legal speed limit is reached
CS>MS	A	Overspeeding, slow down for safety and fuel-saving

d) *Car jerk*: It is the variation of the acceleration during a time, measured in m/s^3. It illustrates a driver's acceleration profile such as forced acceleration that requires more FC to let the car accelerate quickly. As stated in [186], a very robust algorithm can be developed to classify the driver's style with this indicator. It is considered as one of the strongest aggressiveness predictors of crash involvement [187]. This feature was used in several driving profiling studies e.g., [39][131][186]. OBD-II data does not supply a sensor for the jerk. As the longitudinal acceleration 'a' is not provided by the OBD-II interface and the GPS signals, it was estimated from the OBD-II car speed 's' as given in Eq. 6. Having the computed acceleration, then the jerk 'j' was estimated as in Eq. 7. The empirical findings in [188] show that a comfortable (non-jerky) driving pattern occurs when the jerk ranges from -4 m/s^3 to 3 m/s^3. This suggestion was followed

[1] OpenStreetMap http://www.openstreetmap.org

as adopted in [39]. Table 15 illustrates the categorisation for the driver with this indicator. A driver is classified as moderate 'M', if the jerk value ranges in [-4 m/s^3, 3 m/s^3], or as aggressive 'A' if it is not the case. For the second option, the driver is warned with "Avoid forced acceleration" in case the jerk > 3 m/s^3 or by "Avoid sharp deceleration" in case the jerk < -4 m/s^3.

$$a\ (t) = \frac{\mathrm{s(t) - s(t-1)}}{\mathrm{t - (t-1)}} \qquad (6)$$

$$j(t) = \frac{\mathrm{a(t) - a(t-1)}}{\mathrm{t - (t-1)}} \qquad (7)$$

Table 15. Driving classification with car jerk.

Car Jerk (m/s^3)	Class	Recommendation
-4≤jerk≤3	M	-
jerk>3	A	Avoid forced acceleration
jerk<-4	A	Avoid sharp deceleration

4.3.2 Case Study and Discussions

I considered 800 trees, 100 levels and the square root of the number of features to split at each leaf node, as found in my previous work (4.2.1) [5]. As you can see in Table 16, involving engine load as an additional predictor, increases the performance of the model, with a lower MSE (0.82 vs 1.5) and a slightly higher R^2 (0.94 vs 0.896) (refer to Table 11). Fig. 49 shows the fit of the RF model.

Table 16. RF model performance with and without 'engine load'.

Model	RF(-Engine Load)	RF (+Engine Load)
MSE	1.506	0.82
R^2	0.896	0.94

I selected one of the historical tracks with 575 measurements of 71 km with an average time of 50 minutes, recorded in Germany in 2016-02-17, in the time slot 16:00-17:00. This trip was recorded with a Volkswagen Polo 9N 2009, gasoline engine. Fig. 50 presents the geographical visualization of the studied trip, while Fig. 51. shows the analysis done with RStudio[2] for the considered indicators. The picture at the bottom of Fig. 51, visualizes the time evolution of FC predictions by the RF model versus the actual enviroCar estimated one for the studied track.

[2] RStudio https://www.rstudio.com/

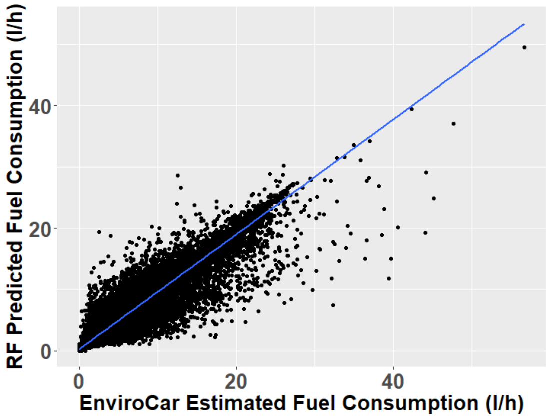


Fig. 49. RF model fit (test set 20% of the data).

It is noticeable, that the trip was mostly driven on a highway since the legal speed limit supplied by OSM, is greater than 100 km/h (except at the start and the end of the timeline) and the driver has a speed, higher than 60 km/h with no big acceleration or deceleration intervals in those cases (apart from the middle of the trace where the speed pattern went lower than 60 km/h on the highway).

There is a clear relationship between the car speed and the engine RPM over time. On the motorway, where the speed is higher than 60 km/h, the values of RPM are higher compared to urban roads at the start and the end of the trip where the OSM speed limit is about 30 km/h. To move the car at such high speeds, the vehicle engine load is higher, and it has to work harder.

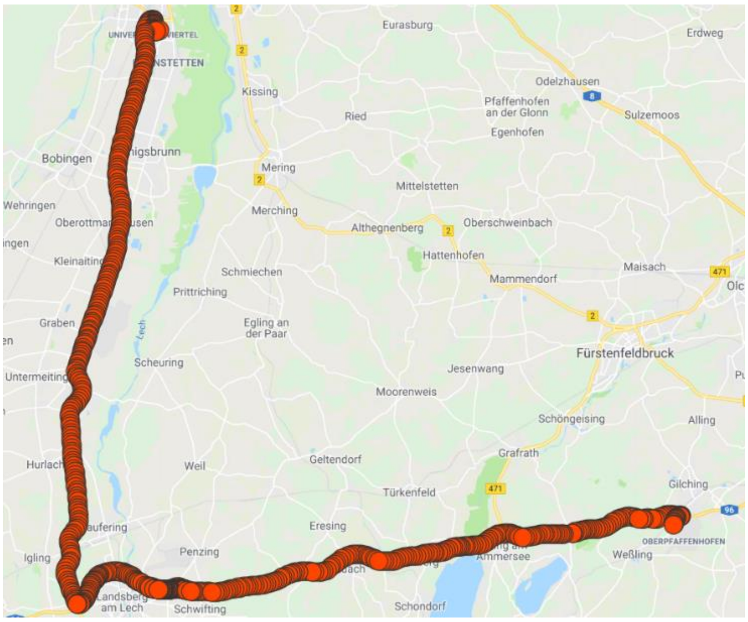


Fig. 50. Geographical visualization of the studied trip.

On the other hand, in the beginning, the middle and the end of the trip, the car moves in a stop-and-go fashion. The reason being the traffic density on an urban road in the beginning and end of the trip, where the vehicle speed adheres to the legal speed limit of less than 60 km/h. This also occurs in the motorway section which has a 120 km/h speed limit in the middle of the trip. This may be due to a traffic jam caused by an accident. In such situation, the acceleration patterns fluctuate more than the case of highspeed travel on a highway, since on an urban road the driver has to accelerate and decelerate more when encountering traffic lights or road crossing to be in idling position, then to accelerate again letting the car to move.

The car jerk clearly depicts (fourth top plot of Fig. 51.) how the driver's acceleration profile changes over time, whether accelerating or decelerating. The jerk peaks in the case of noticeable positive variation in the acceleration or drops for the opposite scenario. For instance, at the beginning of the trip, the jerk is too high since there is a significant variation of acceleration when the driver accelerated in order to move the car from a stable to a running state.

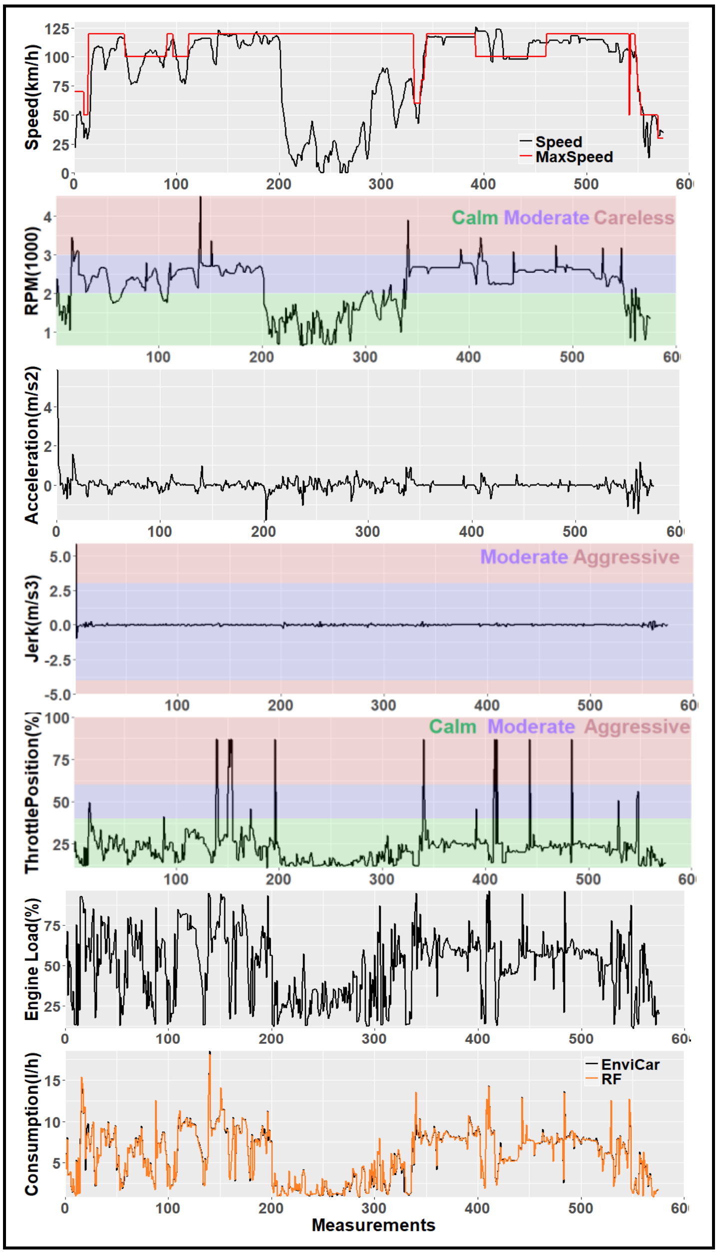


Fig. 51. Car data behaviour along with the trace: speed, RPM, acceleration, car jerk, TPS, engine load and FC (RF vs enviroCar).

Speeding requires a high RPM, leading to a drop in fuel-efficiency. This is clear from the FC timeline of the analysed trip, where more fuel is consumed for the case of high speed. On the motorway (the speed is higher than 60 km/h), the values of RPM are higher compared to urban roads at the start and the end of the trip, where the OSM speed limit is about 70 and 30 km/h respectively. In those cases, the values of engine load are higher, since the engine has to work harder for moving the car at high speeds, which implies more FC.

Further, the FC fluctuates more than the engine RPM on highways. This is caused by engine load variations which is clear from the engine load timeline. This might be due to car configuration changes (e.g., caused by the use of heated seats and demister blowers in cold weather as the trip was recorded at the beginning of February) or changes in the use of car accessories such as playing the entertainment system more loudly. This is the reason for involving the engine load in the FC estimation.

It is clear from the plots of both FC and throttle position (Fig. 51), that higher values of TPS are translated into higher FC values as well, regardless of the impact of other factors on FC. This is true for the opposite direction when a driver releases the accelerator pedal (resulting in a drop in TPS values), the FC values decline at those moments. When there are spikes in TPS, there are peaks in FC pattern too and this is true for the opposite case.

Table 17 provides the driving classification report for each of the considered driving indices in the dynamic recommendations. The trip is considered as having moderate RPM, calm use of the accelerator pedal, with a low percentage of exceeding the permissible speed limit (15%) and with moderate changes in acceleration or deceleration (jerks).

Fig. 52 depicts the achieved fuel-efficiency scores (normalized to 100) for the 111 trips that are recorded with the same car model of the simulated trip (Volkswagen Polo 9N 2009, gasoline engine). It is common to attaint a fuel-efficiency score between 50 and 60. Hence I referred this range to ‘Typical’ class. Therefore, the eco-driving class is ‘Careless’ in case the achieved fuel-efficiency score less than 50/100, and the driving is categorised by fuel ‘Saver’ when the score is over than 60.

Table 17. Example of instantaneous aggressiveness indicators report.

Driving indices	***Classification***		
RPM	C: 32%	M: 65%	CA: 3%
TPS	C: 96%	M: 2%	A: 2%
Car speed	-	M: 85%	A: 15%
Car Jerk	-	M: 99.8%	A: 0.1%

The achieved fuel-efficiency for the studied trip is 0.021 km/l/h, while the maximum attainted fuel-efficiency by 111 tracks (for the same car type, in the same region) is 0.037 km/l/h. Therefore, the reached eco-driving score for the trip is 56.25 over 100, which indicates a typical behaviour.

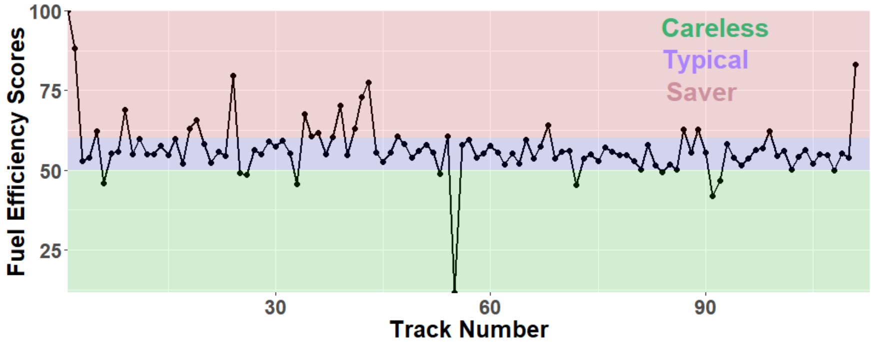


Fig. 52. Fuel-efficiency scores for 111 tracks for car model "Volkswagen Polo 9N 2009", gasoline engine.

4.3.3 Methodology Recap

The presented gaming approach for driving profiling, supplies the driver or player with direct recommendations when no eco-driving events are detectable. This allows the driver to adapt his or her driving to reduce FC and CO2 emissions. The direct feedback is driven from analysing the following inputs; throttle position, RPM, car speed and car jerk (changes in acceleration with respect to time). I relied on the literature in the evaluation of those indicators. These indicators illustrate a driver's aggressiveness resulting in consuming more fuel.

After a drive, the game supplies the drivers with a score between 0 (not efficient) and 100 (efficient, best score) to classify their ecological (such as fuel 'Saver' and 'Typical') in respect to the achieved fuel-efficiency (the most important metric in eco-driving). It can balance a bit the trade-off between the impact of driving style and other influences on the fuel economy. The higher the score, the more the driving is efficient (comparing to other drivers of the same car type), and the more the driver deals carefully with the accelerator pedal (information driven from TPS).

4.4 Methodology 4: Module 2 Towards a Reality-Enhanced Serious Game to Promote Eco-driving in the wild

In this chapter, I propose a new field user performance evaluator methodology of the one presented in the previous methodology in 4.3, for promoting fuel-efficient driving with SGs technologies. The FC estimations of the previous methodologies have been improved in terms of involving more inputs. Yet, I present the application of three ML techniques in term of FC' prediction mainly, support vector machine for regression (SVR), random forest (RF) and artificial neural networks (ANNs). The driving classification after a trip relies on the fuel-efficiency (metric in eco-driving) metric as in 4.3.1 (A).

When inefficient driving attitudes are detectable, the module supplies the driver with the recommendation that is based on the fuzzy rules deduced in 4.1 via sensing and analysing the changes of TPS, RPM and car speed. The recommendation is provided in case the level of the consumed fuel is high.

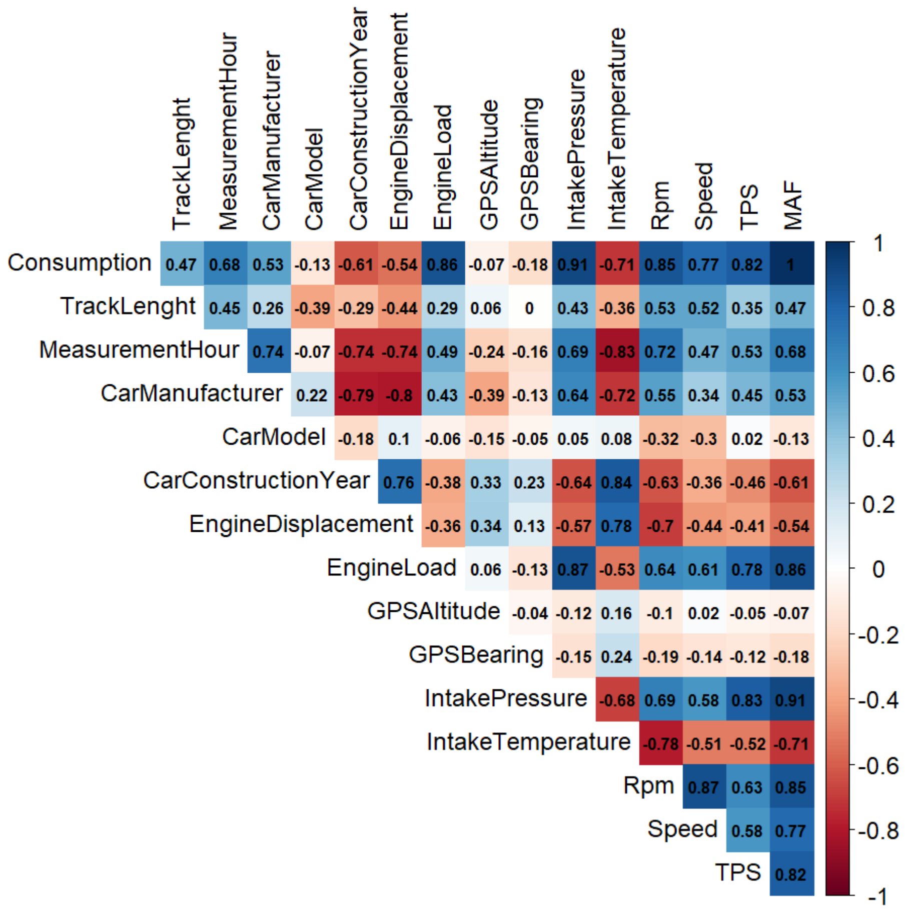


Fig 53. Correlation between enviroCar estimated FC and other variables.

The remaining of this chapter is organised as following; 4.4.1 describes the instant FC estimation including the data analysis for involving the predictors in the FC modelling process; 4.4.2 presents the proposed RT driving feedback; 4.4.3 presents the analysis, assessment and a simulated case study; and ultimately 4.4.4 summarises this methodology.

4.4.1 Instant Fuel Consumption Estimation

I. Data Analysis and Predictors Selection

As a first step, I analyzed the PPMC between the target variable (enviroCar estimated FC) and other variables (including vehicular signals), which is reported in Fig 53. Table 2 explains the colour codes for the PPMC values. A positive sign refers to a positive correlation, while a negative sign indicates a negative correlation. The closer the value between an input and the outcome FC to 1 or -1, the more the variable is likely to be predictive. Apart from MAF (with an obvious PPMC of 1), several signals are highly positively correlated with FC such as intake pressure (0.91), engine load (0.86), RPM (0.85), TPS (0.82) and speed (0.77). Intake-air temperature (-0.71) and car construction year (-0.61), on the other hand, are highly negatively correlated with FC.

I studied 14 predictors that might affect the consumption of fuel as reported in Fig 53. They can be divided into four groups:

1. Variables related to car characteristics: Car manufacturer; car model; car construction year; and engine displacement.
2. Variables read from the internal car's sensors (OBD-II scanner):
   - Engine Load "%": the power that the outside world takes away from the engine. It measures how much air and fuel are sucking into the engine.
   - Speed "km/h": the actual speed of the vehicle shown by the odometer (when there is no readable value from the speed sensor, I have relied on GPS speed): Speeding burns fuel.
   - Intake-air temperature "c": senses the air temperature inside the cylinders incoming to the engine. This information is provided to the ECU for correcting the mixture formation and the ignition to determine the correct amount of fuel needed for optimum performance and economy.
   - Engine speed "u/min" or the number of engine revolutions per minute (RPM): The higher the RPM, the more the fuel is consumed [34]. Optimal RPM value differs between cars (e.g., engine characteristic) and depends on road type (e.g., uphill or downhill).
   - Throttle position sensor (TPS) or accelerator "%": it regulates the air and fuel intake into the engine, making it run slower or faster – the more the throttle pedal is depressed, the more the fuel and air are supplied to the engine and ignited. It is one of the parameters that are controlled directly by drivers, reflecting their habits in dealing with the accelerator pedal.
   - MAF "l/s", described earlier.
   - Intake manifold absolute pressure (MAP) "kPa" used in an internal combustion engine's electronic control system.

3. Computed post-hoc and added on the community's server: FC "l/h"; calculated MAF "g/s" (When OBD adapter delivers no result for MAF, I have relied on this); and tack' length "km" – the travelled distances in kilometres (or the distance from the startup point to the final point) of an individual track.
4. Embedded sensors in smartphone and timestamp data: GPS speed (When OBD-II adapter delivers no result for speed, the GPS speed value is considered in this work) and hour (expressing part of the day).

GPS altitude, GPS bearing, and car model information have been excluded as they are with little correlation with the FC.

II. Fuel Consumption Modeling

I used the above-presented data to develop a new FC model based on ML. I applied three ML techniques in term of FC prediction mainly, SVR, RF and ANNs described in sections 2.4.2, 2.4.3 and 2.4.4 respectively.

I relied on the grid search for tuning both SVR and RF models, while I used the random search to configure the ANN model. Both hyperparameter optimization techniques have been implemented using 'GridSearchCV' [182] and 'RandomizedSearchCV' [181] respectively, of the python sk-learn library. They have been optimized by 10-fold CV for reducing the chance of overfitting. The grid search constructs and evaluates one model for each possible combination of a defined set of values for each hyperparameter. The random search, instead, trains and assesses candidate models by using random combinations of the parameters for a fixed number of iterations to find the best solution for the built model.

1. Support Vector Machine for Regression (SVR)

The 'SVR' function from '*sklearn.svm*' python package [189], has been used to implement the SVR model. I used the radial basis function (RBF) kernel SVR (or Gaussian kernel) as in [146][190], given its power in handling the nonlinearity between the attributes and the labels, and mapping the data samples into a higher dimensional space. In this case, the optimal RBF Kernel values found by the grid search are, 1 for 'C' (the penalty parameter of the error term) and 0.001 for 'gamma' (the Kernel coefficient for RBF that defines how much influence a single training example has).

2. Random forest (RF)

An introduction of the RF approach has been already supplied in 2.4.2. Likewise in 4.2 (4.2.1) and 4.3 (4.3.1), I used the '*RandomForestRegressor*' of '*sklearn.ensemble*' python library [180]. where I optimised the two most important settings, the number of trees in the forest ('n_estimators') and the number of features considered for splitting at each leaf node ('max_features'). The grid search exploitation shows that 800 trees with 9 features are the optimal configurations.

3. Artificial Neural Networks (ANNs)

The ANN model was implemented by the mean of ‘Keras’ python package [191], following the sequential (with successive layers). The best found of random search suggests that

- 400 epochs are adequate to show the entire training dataset to the network during the training phase, with 20 patterns to show to the network before the weights are updated ('batch_size').
- A single hidden layer having 6 neurons provides satisfactory prediction accuracy. Then the followed ANN model’s structure consists of one input layer (receiving data from the twelve inputs), one hidden layer (with six neurons), and one output layer (with one neuron for FC) as illustrated in Fig. 54.
- The 'relu' neuron activation function in the hidden and the output layers.
- 'Adam' as the optimization algorithm with 'normal' for the NN weight initialization.
- Non regularization is needed (‘dropout_rate’ = 0). The dropout percentage (‘dropout_rate’) and the weight constraint (‘weight_constraint’) optimal values of the dropout technique are 0 and 3 respectively. The dropout regularization technique prevents complex co-adaptations on training data in an effort to limit overfitting.

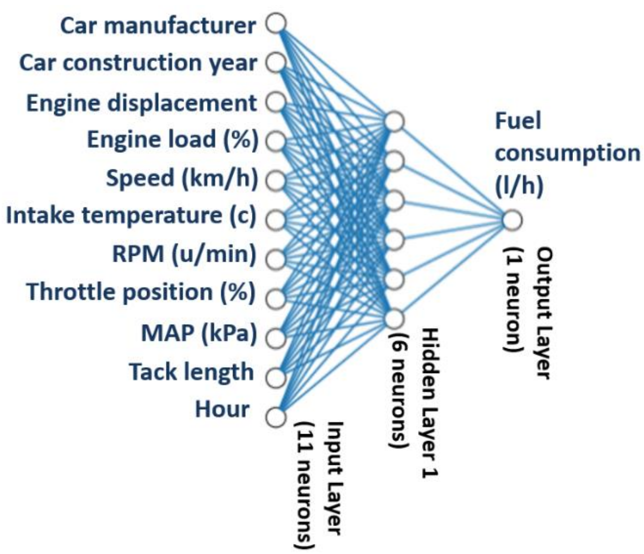


Fig. 54. Representation of the developed ANN model.

4.4.2 Instant Driving Recommendation

The previous section targeted a RT assessment of the driver FC performance. Besides this, it is important also to provide a driver with guidance on how to improve. Table 18 presents the corresponding driving feedback for seven rules extracted from my deduced fuzzy rules in case FC is high ‘H’ or very high

‘VH’ in 4.1 [4] (after studying and observing the 60 possible combinations of the MFs of the variables with the AND operator check FIS4 in 4.1.2 (D)). FL transfers human knowledge and expertise into a mathematical model through if-then rules, that allow providing verbal feedback directly related to the inputs. These features are particularly suited for providing feedback also to non-expert users.

Table 18. Extracted fuzzy rules with relevant proposed feedback in case FC is high or very high (L: low, M: medium, H: high, VH: very high).

	FL Rules (Ri / i=1-7)
R1	if RPM is L & TPS is H & Speed is H then FC is VH
R2	if RPM is L & TPS is H & Speed is VH then FC is H
R3	if RPM is H & TPS is M & Speed is M then FC is H
R4	if RPM is H & TPS is M & Speed is (H or VH) then FC is H
R5	if RPM is H & TPS is H then FC is H
R6	if RPM is VH & TPS is M then FC is H
R7	if RPM is VH & TPS is H then FC is VH
	Corresponding Driving Feedback (Fi / i=1-7)
F1	Downshift the gear and raise the accelerator pedal
F2	Downshift the gear and raise the accelerator pedal
F3	Upshift the gear
F4	Upshift the gear and reduce speed
F5	Upshift the gear and raise the accelerator pedal
F6	Upshift the gear
F7	Upshift the gear

While providing the recommendation, RPM is considered as the strongest FC predictor, followed by TPS and car speed. I deduced this feature importance rank while observing the fuzzy rules of FIS4 in 4.1.2 (D)) [4] and then I confirmed it later in 4.2.2 via RF feature importance interpretation tool for measuring the prediction strength of each variable in [5]. For instance, based upon the 4^{th} rule in Table 18, if a “High” FC is obtained because of a ‘High’ RPM, a ‘Medium’ TPS and a (‘High’ or a ‘Very high’) speed, then the system provides the advice “Upshift the gear and reduce speed” prioritizing the gear shift action to suggest later to decelerate. For the 5^{th} rule (Table 18) – for a ‘High’ level of RPM and similarly, a ‘High’ level for TPS whatever the values of car speed, the FC is ‘High’ – I provide the advice “Upshift the gear and raise the accelerator pedal”.

In providing verbal feedback, I also included an overspeed event detector as a crucial metric to characterize driver safety compliance. Overspeeding events are triggered if the car’s speed is greater than the legal speed limit, which is obtained through a web service access, based on OSM. The driving classification with this indicator is already provided in Table 14.

4.4.3 Results, Case Study and Discussion

The experimental tests were conducted in the same computational environment, on an 8 GB RAM and i7-8550U CPU laptop PC. The developed models were trained on 80% of the available data and tested on the remaining data to judge the quality of the fit. The out-of-sample performance of each model is provided by the statistical metrics, the mean of MSE and R^2. Fig. 55 illustrates the fits of the three models.

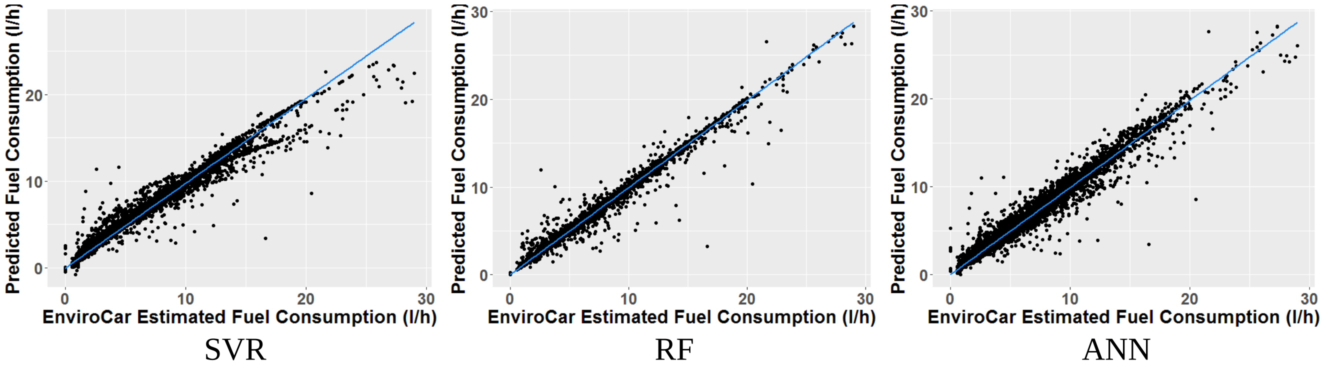


Fig. 55. Comparison of the models (SVR, RF and ANN) fit.

Table 19 provides the likely performance of the models. Results show that the RF slightly outperforms ANN, with a slightly lower MSE (0.02 vs 0.05) and a slightly higher R^2 (0.99 vs. 0.98). Both models outperform the one given by SVR method.

Table 19. Comparison of model performance.

Performance \ Model	SVR	RF	ANN
MSE	0.06	0.02	0.05
R^2	0.98	0.99	0.98
Training time (sec)	9240	720	7020
Response time (mS)	1	27	0.28

Upon the experiments, the RF was the fastest in learning from the data during the training phase, of the order of 720 sec versus 7020 sec and 9240 sec for ANN and SVR respectively (Table 19). Nevertheless, the RF model is the slowest in predicting the FC with a timely response of the order of 27 mS. SVR was the most computationally expensive in fitting the data with the highest training time of the order of 154 mS (Table 19). SVR was the most computationally expensive in fitting the data (of the order of 1 mS), while ANN is the fastest one (of the order of 27 mS).

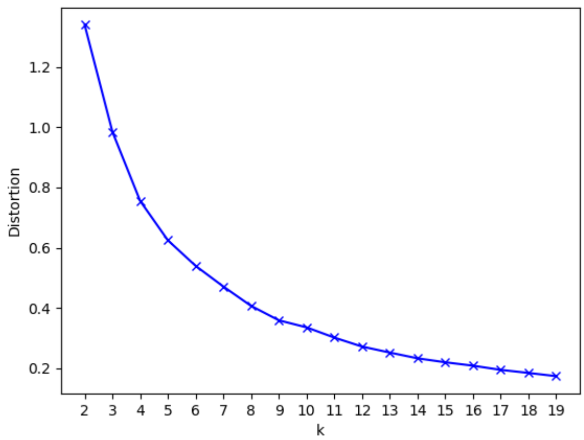


Fig. 56. Elbow chart, a hint for the optimal number of driving categories.

Based on the analysis, it can be concluded that all the three methods achieve excellent results in predicting FC based on the vehicular signals made available by the OBD-II interface and other studied variables related to car characteristics and driving time (Fig 53). RF that is quickly trainable, has the best performance while ANN is the quickest in prediction FC.

Table 20. Centroids of the clusters for k=5 and k=9.

K	Centroids (l/h)
5	1.76, 4.35, 6.93, 9.04, 11.91
9	1.37, 2.65, 4.15, 5.68, 7.17, 8.46, 9.73, 11.45, 14.84

In eco-driving, important feedback to the driver is given by categorization, which is typically provided through the three traffic light colours. Therefore, I relied on the “Elbow” method based on the 1d k-means algorithm to decide the best number of driving style groups. The elbow chart depicted in Fig. 56, shows that the elbow (the point of inflection of the curve) but is clearly above 3 upon the used dataset. If 9 categories may be too many to communicate to the driver, 5 could be a proper trade-off. The corresponding cluster centroids are reported in Table 20.

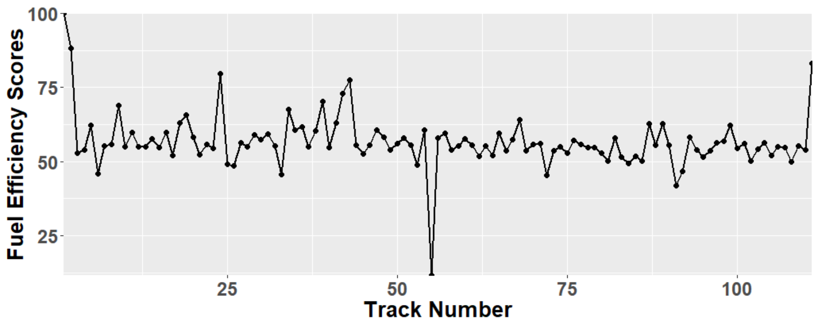


Fig. 57. Fuel efficiency scores for 111 tracks for “Volkswagen Polo 9N 2009” gasoline engine.

In eco-driving profiling, a key metric is given by fuel-efficiency, which relates the amount of consumed fuel by a vehicle and the distance travelled (Eq. 5). I estimate the fuel efficiency for a driver on a trip by dividing the estimated FC by the distance covered. Manufacturers give a fuel efficiency figure for new cars in liters per 100 km under ideal conditions for urban, extra-urban (e.g., higher speeds) and combined (a mixture of the two), which is difficult to achieve. Hence, in order to allow comparisons, I normalize the values by the highest achieved fuel-efficiency with a similar car model as in [185]. This avoids figuring out the maximum fuel-efficiency for every car model.

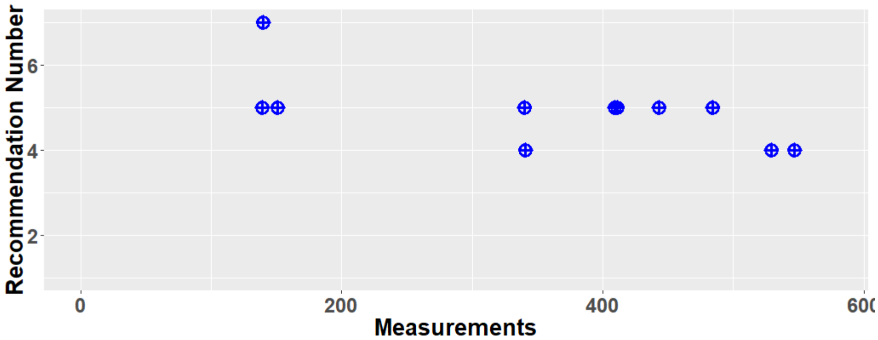


Fig. 58. Driving recommendation timeline for the studied trip respecting the seven rules in Table 18.

The achieved fuel-efficiency for the studied trip is 0.021 km/l/h, while the maximum attainted fuel efficiency by 111 tracks (for the same car type, in the same region) is 0.037 km/l/h. Fig. 57 depicts the achieved fuel efficiency scores (normalized to 100) for the 111 trips relevant to the same car model of the simulated trip (Volkswagen Polo 9N 2009, gasoline engine). It can be noticed that it is common to attain fuel efficiency score between 50 and 60.

Table 21. Example of driving feedback for a measurement of the studied track.

Event detectors	***Values***	***Classifications***
RPM	4530	VH
Car speed/OSM speed	117.36/120 km/h	H, respecting the legal speed limit
TPS	87%	H
******Driving Recommendation******		
Upshift the gear		

Fig. 58 illustrates the timeline of the instant provided verbal feedback along the studied track in 4.3.2, where each blue point represents a recommendation corresponding to one of the seven fuzzy rules in case inefficient driving styles are detectable (Table 18). It is noticeable that rules nr. 5 and 4 are quite frequent, while the others much less. This particularly stresses the importance of proper tuning of the gear shift. Table 21 shows an example of an instant driving recommendation.

4.4.4 Methodology Recap

This chapter investigates three ML techniques, namely, support vector machine for regression (SVR), random forest and artificial neural networks (ANNs), involving eleven predictors that affect the consumption of fuel, including vehicular signals available through the common ODB-II interface for a wide diffusion and a standardized FC module. This involves the case of unreadable or faulty MAF (see section 3.2 (4)). It has been found that ANN is the most powerful technique in the sense of FC response time and accuracy. While the RF model was is the slowest in the FC prediction but the quickest in learning from data (Table 18).

For providing driving recommendations in case inefficient driving attitude is detectable, I relied on the fuzzy rules of the first methodology in 4.1.2, using FL – one of the AI techniques – for keeping the driver aware of fuel economy when inefficient driving patterns are detectable. The rules are provided in case the FC level is high or very high (check MF of FC in 4.1.2). This avoids fixing a threshold to asses the inputs based on the literature where FL can deal with the uncertainty.

5 Conclusion, Discussion and Future Work

Drivers still have rooms to save fuel regardless of technological improvement. This research shows how it is challenging to evaluate/classify driving style, but on the other hand, research shows how motivating advice can help to keep the driver prudent of fuel wasting. There are many measures which have been considered while evaluating driving' patterns. However, the following are the common key measures that have been used in most studies: car speed, dealing with the accelerator pedal (via throttle position signal) and engine speed (revolutions per minute signal).

IoT technologies are spurring typologies of serious games (SGs) (e.g., RESGs shorter for reality-enhanced serious games) that support training directly in the field. This research exploits IoT's potential for SGs in the automotive application domain aiming to improve automotive driver behaviour, as vehicles become ever more powerful IoT platforms [31] and car drivers have significant margins to improve safety and reduce emissions. The contribution of this work consists in identifying two main features of RESG (i.e., instant assessment and recommendation) and in proposing two modules employable as virtual sensors for driver's behaviour assessment and coaching in third-party RESGs to promote fuel-efficient driving.

The first module estimates FC in RT. In this process, I compared the performance of three well-established ML algorithms, mainly support vector machine for regression (SVR), random forest (RF) and artificial neural networks (ANNs). It has been found that the algorithms have similar performance. RF achieves the highest correlation with FC, however, it takes the longest time in prediction (30 mS., still suitable for the RT requirement). FC estimations with good accuracy are needed as continuous sequential determinations of FC can be realised by a game design directly as an ongoing updated score, or as the energy of the player, or to activate bonuses or maluses, or to facilitate reaching a higher gaming level, etc. (e.g., [22][32]). While the second module provides instant recommendations using FL, suggesting actions to be taken when an inefficient driving pattern has been detected.

The algorithms exploit signals available through the On-board diagnostics II (OBD-II) standard interface, which is mandatory on any vehicle. I tested the approaches with data from the enviroCar server site. The data is not calibrated for a specific car model and is recorded in different driving environments, which made the work challenging and robust for real-world conditions. Like the RESG architecture, also the proposed approach to virtual sensor design is general and thus applicable to various application domains other than fuel-efficient driving. Choosing a pluggable SG mechanism for a smartphone as a platform for providing a score for each drive, enables introducing game strategies such as ranking users based upon their achieved efficiency.

The presented models are now available to serious designers for easy integration in a reality-enhanced gaming architecture (e.g., [172]), as the estimated value can then be seamlessly fed in different game logics and treated according to their specificity. It will be interesting to see how game designers will exploit this information within compelling SGs for promoting fuel efficiency. The preprocessed data that I have requested is used now with a new PhD student who is trying to install my models on raspberry pi. Furthermore, this research presents a new platform to parse and save requested enviroCar data from their server, that it will be open-source on GitHub development platform as future work.

5.1.1 System Limitations

While I performed extensive lab tests, real-world tests are needed after installing the models on a smartphone or on a raspberry pi to be connected to the car via Bluetooth connection. This can verify the training and coaching validity of the developed approaches.

An important word of caution is needed concerning users' privacy, as the modules rely on highly sensitive data (e.g., overspeeding, dealing with the accelerator pedal), and provide information that by no means should be misused [122].

Since fuel consumption is a dimension of the driving style, further research is needed for keeping into account and assessing other relevant aspects, such as safety. Safety should be considered carefully in eco-driving gaming applications, since unsafe fuel-saving comportments will be reinforced any time fuel-saving in general [57], e.g., letting the vehicle to move not under the driver's control with coasting function of hypermiling while driving (2.2.5). Also, eco-driving systems have to be non-intrusive and not highly demanding user attention for driver's distraction and irritation avoidance. Yet, drivers should keep paying more attention to the road ahead focusing on their surroundings (pedestrians and other vehicles) rather than to the phone or the screen of the car where the driving game is displayed. Therefore, a well studied and tested gaming design should be considered avoiding visual distracting with the game design as discussed in section 2.3.5.

Ultimately, it is important to consider road type while evaluating driving' pattern. I have involved the speed limit requested by OpenStreetMap (4.3) to bypass this challenge. However other factors can be involved as mentioned later in 5.1.2.

5.1.2 Future Work

The main focus of future work could be summarised with the following points:

1. Validate the effectiveness of the proposed driving profiling methodology with RT testing: The developed models are now available to designers for integration in new SGs design.

   a. Hence multiple drivers can be evaluated under different driving conditions (e.g., riskier or normal).
   b. This shows the provision of providing feedback while driving in terms of improving fuel economy and of driving distraction concern that menace the drivers ‘safety.
2. The two different approaches of providing feedback in 4.3 and in 4.4 (referring to the deduced rules of 2.4.1) can be evaluated with real driving tests.
3. Contribute with the Java software for requesting, parsing and saving locally enviroCar data to GitHub public platform for helping other researchers to exploit this data.
4. Exploit how smartphone sensors can predict the consumption of fuel (avoiding the use of OBD-II adapters) and how accurate are those estimations.
5. Considering other types of engines than gasoline, such as the electric or the hybrid engines that are in future fuel-efficient vehicles.
6. Expend the data that I used to cover move the driving influences taken into account that external influences can help more in understanding the driving style: since driving environment affects hardly a driving pattern, the data from enviroCar that I have used can be linked to other sources of data such as road trajectories, road inclination for the state of the road or air pollution data. This involves yet, the request of weather data that I have started as mentioned in 3.3.
7. Work on other defuzzification techniques of the developed fuzzy logic model in 4.1 and check how they influence the outcomes.
8. A machine learning approach could be tested to generate fuzzy rules automatically rather than data analysis that I have done (4.1.2).

Publications

➢ Journal

1. Rana Massoud, Stefan Poslad, Francesco Bellotti, Riccardo Berta, Kamyar Mehran, and Alessandro De Gloria. A fuzzy logic module to estimate driver's fuel consumption for reality-enhanced serious games. International Journal of Serious Games, 5(4):45–62, 2018.
2. Rana Massoud, Francesco Bellotti, Stefan Poslad, Riccardo Berta, and Alessandro De Gloria. IoT Sensing for Reality-Enhanced Serious Games, a Fuel-Efficient Drive Use Case. IEEE Transactions on Games (ToG). Submitted in Dec 2019.

➢ Conferences Proceedings

1. Rana Massoud, Francesco Bellotti, Stefan Poslad, Riccardo Berta, and Alessandro De Gloria. Exploring Fuzzy Logic and Random Forest for Car Drivers' Fuel Consumption Estimation in IoT-Enabled Serious Games. IEEE International Symposium on Autonomous Decentralized Systems (ISADS) 2019, in press.
2. Rana Massoud, Francesco Bellotti, Stefan Poslad, Riccardo Berta, and Alessandro De Gloria. Eco-driving Profiling and Behavioural Shifts Using IoT Vehicular Sensors Combined with Serious Games Eco-driving Profiling and Behavioural Shifts Using IoT Vehicular Sensors Combined with Serious Games. IEEE Conference On Games (COG) 2019, in press.
3. Rana Massoud, Francesco Bellotti, Stefan Poslad, Riccardo Berta, and Alessandro De Gloria. Towards a Reality-Enhanced Serious Game to Promote Eco-Driving in the Wild. GALA conf 2019: Games and Learning Alliance conference, in press.

Educational and training activities

	Attended Courses			
	Title (duration)	**Date**	**Instructor**	**Organizer**
1	Data Fusion and Bayesian Interaction Modeling for Cognitive Ambient Intelligence (15 h)	20-24 Mar. 2017	Carlo Regazzoni	University of Genova, Italy
2	Industrial Analytics: Theory and Practice of Learning from Data (15 h)	3-7 Jul. 2017	Davide Anguita & Luca Oneto	University of Genova, Italy
3	R Programming A-Z™: R For Data Science With Real Exercises! (10.5 h)	Sep. 2017	Kirill Eremenko	Udemy, online
4	R Programming: Advanced Analytics In R For Data Science (6 h)	Sep. 2017	Kirill Eremenko	Udemy, online
5	Machine Learning A-Z™: Hands-On Python & R In Data Science (41 h)	Feb. 2018	Kirill Eremenko	Udemy, online
6	Introduction to R (7.5 h)	Mar. 2018	Florian Weiler	University of Bamberg, Germany
7	Panel Data Analysis (15 h)	Mar. 2018	Henriette Engelhardt-Wölfler	University of Bamberg, Germany
8	Neural Networks and Deep Learning (15 h)	25-29 May 2018	A. Suárez, J.R. Dorronsoro, A. Barbero	University Politécnica de Madrid Spain
9	Bayesian Networks (15 h)	25-29 May 2018	Pedro Larrañaga & Concha Bielza & Bojan Mihaljevic	University Politécnica de Madrid Spain
10	English Pronunciation Skills (15 h)	Oct. 2018		QMUL, UK
11	LaTeX Tutorial for Beginners (3 h)	01 Nov. 2018	Augusto Nembrini da Rocha	QMUL, UK
12	LaTeX For Everyone and Everything (2.5 h)	29 Jan. 2019	Mohammad Nauman	Udemy, online

	Attended Seminars/Lectures (1 hour)			
	Title	**Date**	**Speaker**	**Organizer**
1	How does health feel? Exploring use of materials in technology led behavior change for wellbeing	21 Jun. 2018	Marion Lean	QMUL, UK
2	Operational Research today: strengths and opportunities	13 June 2018	Ruth Kaufman	QMUL, UK
3	Learning the degree of appearance similarity between two images for person Re-Identification	03 Jul. 2018	Maria Jose	QMUL, UK
4	Gamification at University: not Child's Play	28 Jan. 2019	Sergi Robles	QMUL, UK
5	Defining and measuring challenge as player experience in video games	3 Jul. 2019	Alena Denisova	QMUL, UK

	Attended Winter/Summer Schools		
	Name	**Date**	**Organizer**
1	ECPR Winter School in Methods and Techniques	02-09 Mar. 2018	University of Bamberg, Germany
2	UPM Advanced Statistics and Data Mining	25-29 June 2018	University Politécnica de Madrid Spain
3	Intelligent Sensing	29-30 Aug. 2018	CIS, QMUL, UK

	Attended Workshops			
	Name	**Date**	**Provided by**	**Organizer**
1	User Experience of Serious Games workshop	May 2017	Panagiotis Zacharias (Open University of Cyprus)	Univ. of Genova, Italy
2	Introducing WebVRL Creating Immersive 3d Worlds	13 Sept. 2018	Tara Collingwoode-Williams, Rob Homewood	QMUL, UK
3	Research Writing Workshops (15 h)	03-31 Oct. 2018	Weronika Fernando	QMUL, UK
4	Making a Poster Presentation (3 h)	07 Nov. 2018	Andrea Cox	QMUL, UK
5	Working With Your Supervisor (3 h)	16 Nov. 2018	Fryni Panayidou	QMUL, UK
6	LinkedIn secrets - generating job options (1.5 h)	07 Dec. 2018	Andrea Cox	QMUL, UK
7	Strategic job hunting (UK and abroad) (2 h)	11 Dec. 2018	Andrea Cox	QMUL, UK
8	Indoor People Tracking and Activities of Daily Living Recognition Using Smartphones (2h)	12 Feb. 2019	Bang Wu, Guangyuan Zhang	QMUL, UK

	Attended Conferences	
	Name	**Organizer**
1	Intelligent Games and Game Intelligence (IGGI), 12-13 Sep. 2018	QMUL, UK
2	The 14th IEEE International Symposium on Autonomous Decentralized Systems (ISADS), 8-10 Apr. 2019	HU University of Applied Sciences Utrecht, Netherlands
3	IEEE CONFERENCE ON GAMES (COG), 20-23 Aug. 2019	QMUL, UK

Bibliography

[1] Hans Jakob Walnum and Morten Simonsen. Does driving behavior matter? an analysis of fuel consumption data from heavy-duty trucks.Transportation research part D: transport and environment, 36:107–120, 2015.

[2] Rosolino Vaiana, Teresa Iuele, Vittorio Astarita, et al. Driving behavior and traffic safety: An acceleration-based safety evaluation procedure for smartphones. Modern Applied Science, 8(1):88, 2014.

[3] Michael Sivak and Brandon Schoettle. Eco-driving: Strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. Transport Policy, 22:96–99, 2012.

[4] Rana Massoud, Stefan Poslad, Francesco Bellotti, et al. A fuzzy logic module to estimate a driver's fuel consumption for reality-enhanced serious games. International Journal of Serious Games, 5(4):45–62, 2018.

[5] Rana Massoud, Francesco Bellotti, Stefan Poslad, et al. Exploring Fuzzy Logic and Random Forest for Car Drivers' Fuel Consumption Estimation in IoT-Enabled Serious Games. IEEE International Symposium on Autonomous Decentralized Systems (ISADS), 2019, in press.

[6] Jack N Barkenbus. Eco-driving: An overlooked climate change initiative. Energy Policy, 38(2):762–769, 2010.

[7] Jérôme Barbé, Guy Boy, et al. On-board system design to optimize energy management. In Proceedings of the European Annual Conference on Human Decision-Making and Manual Control (EAM 2006), Valenciennes, France, pages 27–29, 2006.

[8] Olavi H Koskinen. Improving vehicle fuel economy and reducing emissions by driving technique. In 15th World Congress on Intelligent Transport Systems and ITS America's 2008 Annual MeetingITS AmericaERTICOITSJapanTransCore, 2008.

[9] Joeri Van Mierlo, Gaston Maggetto, Erik Van de Burgwal, et al. Driving style and traffic measures-influence on vehicle emissions and fuel consumption. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 218(1):43–50, 2004.

[10] V Corcoba Magana and Mario Munoz-Organero. GAFU: Using a gamification tool to save fuel. Intelligent Transportation Systems Magazine, 7(2):58–70. IEEE, 2015.

[11] European Environment Agency. Transport indicators tracking progress towards environmental targets in Europe. TERM 2011. Retrieved from https://www.eea.europa.eu/publications/foundations-for-greener-transport on 2019-10-10.

[12] Xiaohua Zhao, Yiping Wu, Jian Rong, et al. Development of a driving simulator based eco-driving support system. Transportation Research Part C: Emerging Technologies, 58:631–641, 2015.

[13] Wenshuo Wang, Junqiang Xi, and Huiyan Chen. Modeling and recognizing driver behavior based on driving data: A survey. Mathematical Problems in Engineering, 2014.

[14] Clara Marina Martinez, Mira Heucke, Fei-Yue Wang, et al. Driving style recognition for intelligent vehicle control and advanced driver assistance: A survey. Transactions on Intelligent Transportation Systems, 19(3):666–676. IEEE, 2017.

[15] Narelle Haworth and Mark Symmons. Driving to reduce fuel consumption and improve road safety. Monash University Accident Research Centre. http://www.rs conference. com/pdf/RS010036. pdf, 2001.

[16] Mark S Young, Stewart A Birrell, and Neville A Stanton. Safe driving in a green world: A review of driver performance benchmarks and technologies to support 'smart'driving. Applied Ergonomics, 42(4):533–539, 2011.

[17] Guy H Walker, Neville A Stanton, and Mark S Young. Hierarchical task analysis of driving: A new research tool. Contemporary ergonomics, pages 435–440, 2001.

[18] Johannes Tulusan, Lito Soi, Johannes Paefgen, et al. Eco-efficient feedback technologies: Which eco-feedback types prefer drivers most? In international symposium on a world of wireless, mobile and multimedia networks, pages 1–8. IEEE, 2011.

[19] Richard Van Eck. Digital game-based learning: It's not just the digital natives who are restless. EDUCAUSE review, 41(2):16, 2006.

[20] Ioanna Iacovides. Exploring the link between player involvement and learning within digital games. In Proceedings of the 23rd British HCI group annual conference on people and computers: Celebrating people and technology, pages 29–34. British Computer Society, 2009.

[21] Jan Dirk Fijnheer and Herre Van Oostendorp. Steps to design a household energy game. In International conference on games and learning alliance, pages 12–22. Springer, 2015.

[22] Richardos Drakoulis, Francesco Bellotti, Ioannis Bakas, et al. A gamified flexible transportation service for on-demand public transport. Transactions on Intelligent Transportation Systems, 19(3):921–933. IEEE, 2018.

[23] Thomas Dietz, Gerald T Gardner, Jonathan Gilligan, et al. Household actions can provide a behavioral wedge to rapidly reduce US carbon emissions. Proceedings of the National Academy of Sciences, 106(44):18452–18456, 2009.

[24] Francesco Bellotti, Riccardo Berta, Alessandro De Gloria, et al. Vegame: exploring art and history in venice. Computer, , (9):48–55, 2003.

[25] Paulo HL Rettore, André B Campolina, Leandro A Villas, et al. A method of eco-driving based on intra-vehicular sensor data. In Symposium on Computers and Communications (ISCC), pages1122–1127. IEEE, 2017.

[26] Arne Bröring, Albert Remke, Christoph Stasch, et al. EnviroCar: A Citizen Science Platform for Analyzing and Mapping Crowd-Sourced Car Sensor Data. Transactions in GIS, 19(3):362–376, 2015.

[27] João C Ferreira, José de Almeida, and Alberto Rodrigues da Silva. The impact of driving styles on fuel consumption: A data-warehouse-and-data-mining-based discovery process. Transactions on Intelligent Transportation Systems, 16(5):2653–2662. IEEE, 2015.
[28] Faran Baig, Muhammad Waseem Ashraf, Muhammad Imran, et al. Fuzzy Logic based Fuel Consumption System. Bahria University Journal of Information & Communication Technology, 5(1):31, 2012.
[29] Jason Bordoff and Pascal Noel. Pay-as-You-Drive Auto Insurance. Issues of the Day: 100 Commentaries on Climate, Energy, the Environment, Transportation, and Public Health Policy, 150, 2010.
[30] Jan Willem Bolderdijk, Jasper Knockaert, EM Steg, et al. Effects of Pay-As-You-Drive vehicle insurance on young drivers' speed choice: Results of a Dutch field experiment. Accident Analysis & Prevention, 43(3):1181–1186, 2011.
[31] Richard Kirk. Cars of the future: the Internet of Things in the automotive industry. Network Security, 2015(9):16–18, 2015.
[32] Gautam Ravindra Dange, Pratheep Kumar Paranthaman, Francesco Bellotti, et al. Deployment of serious gaming approach for safe and sustainable mobility. In Intelligent Vehicles Symposium (IV). IEEE, 2017.
[33] Ron Cogan. Honda's Eco Assist Makes You a More Efficient Driver (2014). Green Car Journal. Retrieved from https://greencarjournal.com/top-stories/hondas-eco-assist-makes-you-a-more-efficient-driver/ on 2019-10-27.
[34] Derick A Johnson and Mohan M Trivedi. Driving style recognition using a smartphone as a sensor platform. In 14th International Conference on Intelligent Transportation Systems (ITSC), pages 1609–1615. IEEE, 2011.
[35] Sunitha Godavarty, Sam Broyles, and Micheal Parten. Interfacing to the on-board diagnostic system. In Vehicular Technology Conference. 52nd Vehicular Technology Conference (Cat. No.00CH37152), volume 4. IEEE VTS Fall VTC 2000.
[36] Stefan Diewald, Andreas Möller, Luis Roalter, et al. Gameful design in the automotive domain: review, outlook and challenges. In Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, pages 262–265. ACM, 2013.
[37] Fengzhong Qu, Fei-Yue Wang, and Liuqing Yang. Intelligent transportation spaces: vehicles, traffic, communications, and beyond. IEEE Communications Magazine, 48(11):136–142, 2010.
[38] Bruce Wallace, Mike Rockwood, Rafik Goubran, et al. Measurement of vehicle acceleration in studies of older drivers from GPS position and OBDII velocity sensors. In Medical Measurements and Applications (MeMeA) International Symposium on, pages 391–396. IEEE, 2015.
[39] Oussama Derbel and Rene Jr Landry. Driving style assessment based on the GPS data and fuzzy inference systems. In Systems, Signals & Devices (SSD), 12th International Multi-Conference on, pages 1–8. IEEE, 2015.
[40] Julien Monteil, Niall OHara, Vinny Cahill, et al. Real-time estimation of drivers' behaviour. In Intelligent Transportation Systems (ITSC), 18th International Conference on, pages 2046–2052. IEEE, 2015.
[41] Chris M Atkinson, Theresa W Long, and Emil L Hanzevack. Virtual sensing: a neural network-based intelligent performance and emissions prediction system for on-board diagnostics and engine control. Technical report, SAE Technical Paper, 1998.
[42] Latha P S, Menaga N, Manjunath T K. Efficient Estimation of Fuel Consumption Using Fuzzy Logic. International Journal of Engineering Trends and Technology (IJETT), Vol.16, Oct 2014.
[43] R Luther and Peter Baas. Eco-driving scoping study. 2011.
[44] Hussam Achour, JG Carton, and Abdul Ghani Olabi. Estimating vehicle emissions from road transport, case study: Dublin City. Applied energy, 88(5):1957–1964, 2011.
[45] OBD-II PIDs Wiki, https://en.wikipedia.org/wiki/OBD-II_PIDs on 17 July 19.
[46] Jyong Lin, Shih-Chang Chen, Yu-Tsen Shih, et al. A study on remote on-line diagnostic system for vehicles by integrating the technology of OBD, GPS, and 3G. World Academy of Science, Engineering and Technology, 56:435–441, 2009.
[47] There Wasn't a Light on the Dash: Malfunction Indicator Lamp Myths. Retrieved from https://astech.com/news/there-wasnt-a-light-on-the-dash-malfunction-indicator-lamp-myths/ on 2016-11-20.
[48] Diagnostic Trouble Codes for OBD II (SAE J2012). Retrieved from http://www.fastfieros.com/tech/diagnostic_trouble_codes_for_obdii.htm on 2019-07-16.
[49] The OBD II Home. Retrieved from http://www.obdii.com/background.html on 2019-07-14.
[50] OBDII Car Diagnostic Reader - Bluetooth to Android, PC. Retrieved from http://www.elite-electronix.com/car-accessories/obdii-car-diagnostic-reader-bluetooth-to-android-pc on 2019-08-10.
[51] OBD2 explained, CSS. Retrieved from https://www.csselectronics.com/screen/page/simple-intro-obd2-explained/language/en on 2019-07-17.
[52] OBD-II 16-Pin Data Link Connector DLC. Retrieved from https://www.julesbartow.com/Vehicles/Electronics/Index.php on 2016-11-25.
[53] Marco Di Natale, Haibo Zeng, Paolo Giusto, et al. Understanding and using the controller area network communication protocol: theory and practice. Springer Science & Business Media, 2012.
[54] Kiril Mucevski, Automotive CAN Bus System Explained. Retrieved from https://www.linkedin.com/pulse/automotive-can-bus-system-explained-kiril-mucevski/ on 2017-11-01.

[55] Can bus (controller area network). Retrieved from https://autoditex.com/page/can-bus--controller-area-network-34-1.html on 2019-11-11.
[56] Matthew Barth and Kanok Boriboonsomsin. Energy and emissions impacts of a freeway-based dynamic eco-driving system. Transportation Research Part D: Transport and Environment, 14(6):400–410, 2009.
[57] Angela Sanguinetti, Ken Kurani, and Jamie Davies. The many reasons your mileage may vary: Toward a unifying typology of eco-driving behaviors. Transportation Research Part D: Transport and Environment, 52:73–84, 2017.
[58] How to Hypermile, "WikiHow". Retrieved from https://www.wikihow.com/Hypermile on 2019-01-12.
[59] Coasting, "drivingtesttips.biz". Retrieved from https://www.drivingtesttips.biz/what-is-coasting.html on 2018-01-12.
[60] Frank Primerano, Michael AP Taylor, Ladda Pitaksringkarn, et al. Defining and understanding trip chaining behaviour. Transportation, 35(1):55–72, 2008.
[61] Carpool. Retrieved from en.wikipedia.org/wiki/Carpool on 2018-01-12.
[62] Cruise control. Retrieve from https://en.wikipedia.org/wiki/Cruise_control on 2018-01-09.
[63] Eva Ericsson. Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. Transportation Research Part D: Transport and Environment, 6(5):325–345, 2001.
[64] Lee Chapman. Transport and climate change: a review. Journal of transport geography, 15(5):354–367, 2007.
[65] Xabiel G Pañeda, Roberto Garcia, Gabriel Diaz, et al. Formal characterization of an efficient driving evaluation process for companies of the transport sector. Transportation Research Part A: Policy and Practice,94:431–445, 2016.
[66] Stewart Birrell, James Taylor, Andrew Mc Gordon, et al. Analysis of three independent real-world driving studies: A data driven and expert analysis approach to determining parameters affecting fuel economy. Transportation research part D: transport and environment, 33:74–86, 2014.
[67] Paulo HL Rettore, André B Campolina, Leandro A Villas, et al. Identifying relationships in vehicular sensor data: A case study and characterization. In Proceedings of the 6th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications, pages 33–40. ACM, 2016.
[68] Leverson Boodlal and Kun-Hung Chiang. Study of the Impact of a Telematics System on Safe and Fuel-efficient Driving in Trucks. Technical report, 2014.
[69] Habib Chaari, Eric Ballot, et al. Fuel consumption assessment in delivery tours to develop eco driving behaviour. In European Transport Conference, 2012.
[70] CGI, "Modeling the Relation Between Driving Behavior and Fuel Consumption," pp. 1–10, 2014.
[71] Ian Jeffreys, Genevieve Graves, and Michael Roth. Evaluation of eco-driving training for vehicle fuel use and emission reduction: A case study in Australia. Transportation Research Part D: Transport and Environment, 60:85–91, 2018.
[72] Michelle Rutty, Lindsay Matthews, Daniel Scott, et al. Using vehicle monitoring technology and eco-driver training to reduce fuel use and emissions in tourism: a ski resort case study. Journal of Sustainable Tourism, 22(5):787–800, 2014.
[73] Elliot Martin, Nelson Nelson D. Chan, and Susan Shaheen. How public education on ecodriving can reduce both fuel use and greenhouse gas emissions. Transp. Res. Rec., 2287:163–173, 2012.
[74] Slim Azzi, Gilles Reymond, Frédéric Mérienne, et al. Eco-driving performance assessment with in-car visual and haptic feedback assistance. Journal of Computing and Information Science in Engineering, 11(4):041005, 2011.
[75] Jeffrey Gonder, Matthew Earleywine, and Witt Sparks. Final report on the fuel saving effectiveness of various driver feedback approaches. Technical report, National Renewable Energy Lab. (NREL), Golden, CO (UnitedStates), 2011.
[76] Bryan Reimer, Lisa A D'Ambrosio, Joseph F Coughlin, et al. Using self-reported data to assess the validity of driving simulation data. Behavior research methods, 38(2):314–324, 2006.
[77] Ying Wang, Bruce Mehler, Bryan Reimer, et al. The validity of driving simulation for assessing differences between in-vehicle informational interfaces: A comparison with field testing. Ergonomics, 53(3):404–420, 2010.
[78] K Satou, R Shitamatsu, M Sugimoto, et al. Development of the on-board eco-driving support system. Международный научный журнал Альтернативная энергетика и экология, (9):35–40, 2010.
[79] Kanok Boriboonsomsin, Matthew J Barth, and Alexander Vu. Evaluation of driving behavior and attitude toward eco-driving: Southern California limited case study. Technical report, 2011.
[80] Elliot W Martin, Kanok Boriboonsomsin, Nelson D Chan, et al. Dynamic ecodriving in northern california: A study of survey and vehicle operations data from an ecodriving feedback device. In 92nd Annual Meeting of the Transportation Research Board, Washington, DC, 2013.
[81] Heewon Lee, Woohun Lee, and Youn-Kyung Lim. The effect of eco-driving system towards sustainable driving behavior. InCHI'10 Extended Abstracts on Human Factors in Computing Systems, pages 4255–4260. ACM, 2010.
[82] Zhang Xuemin, Ge Yunshan, Zhang Yu, et al. Research on fuel consumption measurement using carbon balance method. Vehicle Engine, 3:56–58, 2005.
[83] L Xu. Comparative analysis and optimization of car-following models for emission estimation. PhD thesis, Master thesis. Beijing Jiaotong University, China. Google Scholar, 2012.
[84] Ebru Dogan, Linda Steg, and Patricia Delhomme. The influence of multiple goals on driving behavior: The case of safety, time saving, and fuel saving. Accident Analysis & Prevention, 43(5):1635–1643, 2011.
[85] Gys Meiring and Hermanus Myburgh. A review of intelligent driving style analysis systems and related artificial intelligence algorithms. Sensors, 15(12):30653–30682, 2015.

[86] Md Saniul Alam and Aonghus McNabola. A critical review and assessment of Eco-Driving policy & technology: Benefits & limitations. Transport Policy, 35:42–49, 2014.
[87] What Is Coasting? “drive Johnsons”. Retrieved from https://www.drivejohnsons.co.uk/learning-centre/driving-dictionary/coasting/ on 2019-04-29.
[88] Joël Lelong. Vehicle noise emission: Evaluation of tyre/road and motor-noise contributions. In PROCEEDINGS OF INTERNOISE 99-THE 1999INTERNATIONAL CONGRESS ON NOISE CONTROL ENGINEERINGHELD IN FORT LAUDERDALE, FLORIDA, USA, 6-8 DECEMBER 1999, VOLUME 1, 1999.
[89] Sandberg U 2001. Noise Emissions of Road Vehicles, Effect of Regulations. Stockholm, Sweden, Swedish National Road and Transport Research Institute, International Institute of Noise Control Engineering, IINCE Working Party on Noise Emissions of Road Vehicles Final Report No. 01-1
[90] Deterding Sebastian, Miguel Sicart, Lennart Nacke, et al. Gamification. using game-design elements in non-gaming contexts. In CHI’11 extended abstracts on human factors in computing systems, pages 2425–2428. ACM, 2011.
[91] Francesco Bellotti, Riccardo Berta, Alessandro De Gloria, et al. A gamified short course for promoting entrepreneurship among ICT engineering students. In 13th International Conference on Advanced Learning Technologies, pages 31–32. IEEE, 2013.
[92] JH Jung, Christoph Schneider, and Joseph Valacich. Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. Management science, 56(4):724–742, 2010.
[93] Kai Erenli. The impact of gamification: A recommendation of scenarios for education. In 15th International Conference on Interactive Collaborative Learning (ICL), pages 1–8. IEEE, 2012.
[94] IBM Innov8 game https://www.ibm.com/developerworks/library/ws-bpm-innov8/ on 22 July 19.
[95] Danae Vara, Enric Macias, Sergio Gracia, et al. Meeco: Gamifying ecology through a social networking platform. In International Conference on Multimedia and Expo, pages 1–6. IEEE, 2011.
[96] Rémi Cnali. Game Design for Eco Driving (2017). Megamification. Retrieved from http://www.megamification.com/game-design-for-eco-driving/ on 2019-10-27.
[97] In Efforts To Be Green, Ford Receives Patents For SmartGauge With EcoGuide (2017). TechCrunch. Retrieved from https://techcrunch.com/2011/03/17/in-efforts-to-be-green-ford-receives-patents-for-smartgauge-with-ecoguide/ on 2019-10-27.
[98] Francesco Bellotti, Riccardo Berta, and Alessandro De Gloria. Designing effective serious games: opportunities and challenges for research. International Journal of Emerging Technologies in Learning (iJET), 5(2010), 2010.
[99] Dragon Box Elements. Retrieved from https://dragonbox.com/products/elements on 2019-07-19.
[100] Stefan Poslad, Athen Ma, Zhenchen Wang, et al. Using a smart city IoT to incentivise and target shifts in mobility behaviour—Is it a piece of pie? Sensors, 15(6):13069–13096, 2015.
[101] Ala Al-Fuqaha, Mohsen Guizani, Mehdi Mohammadi, et al. Internet of things: A survey on enabling technologies, protocols, and applications. Communications surveys & tutorials, 17(4):2347–2376. IEEE, 2015.
[102] Helmut Prendinger, João Oliveira, João Catarino, et al. ICO2: A networked game for collecting large-scale eco-driving behavior data. Internet Computing, 18(3):28–35. IEEE, 2014.
[103] Foong Li Law, Zarinah Mohd Kasirun, and Chun Kiat Gan. Gamification towards sustainable mobile application. In Malaysian Conference in Software Engineering, pages 349–353. IEEE, 2011.
[104] Thomas Olutoyin Oshin, Stefan Poslad, and Zelun Zhang. Energy-efficient real-time human mobility state classification using smartphones. Transactions on Computers, 64(6):1680–1693. IEEE, 2014.
[105] Anton Gustafsson, Cecilia Katzeff, and Magnus Bang. Evaluation of a pervasive game for domestic energy engagement among teenagers. Computers in Entertainment (CIE), 7(4):54, 2009.
[106] Sebastian Deterding, Dan Dixon, Rilla Khaled, et al. From game design elements to gamefulness: defining gamification. In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, pages 9–15. ACM, 2011.
[107] Martijn Kors, Erik D Van der Spek, and Ben AM Schouten. A Foundation for the Persuasive Gameplay Experience. In FDG, 2015.
[108] Carsten Magerkurth, Adrian David Cheok, Regan L Mandryk, et al. Pervasive games: bringing computer entertainment back to the real world. Computers in Entertainment (CIE3(3):4–4, 2005.
[109] U. Ritterfeld, Ute Ritterfeld, Michael Cody, et al. Serious games: Mechanisms and effects. Routledge, 2009.
[110] Inseok Hwang, Youngki Lee, Taiwoo Park, et al. Toward a mobile platform for pervasive games. In Proceedings of the first ACM international workshop on Mobile gaming, pages 19–24. ACM, 2012.
[111] Leif Oppermann and Michaela Slussareff. Pervasive games. In Entertainment Computing and Serious Games, pages 475–520. Springer, 2016.
[112] Petra Sundström, David Wilfinger, Alexander Meschtscherjakov, et al. The car as an arena for gaming. In Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services companion, pages 233–236. ACM, 2012.
[113] Haibo Mei, Stefan Poslad, and Shuang Du. A game-theory based incentive framework for an intelligent traffic system as part of a smart city initiative. Sensors, 17(12):2874, 2017.

[114] Sven Krome, Steffen P Walz, Stefan Greuter, et al. Exploring game ideas for stresslessness in the automotive domain. In Proceedings of the 2014 Conference on Interactive Entertainment, pages 1–3. ACM, 2014.
[115] Rod McCall and Vincent Koenig. Gaming concepts and incentives to change driver behaviour. In Ad Hoc Networking Workshop (Med-Hoc-Net), The 11th Annual Mediterranean, pages 146–151. IEEE, 2012.
[116] Fabius Steinberger, Ronald Schroeter, and Christopher N Watling. From road distraction to safe driving: Evaluating the effects of boredom and gamification on driving behaviour, physiological arousal, and subjective experience. Computers in Human Behavior, 75:714–726, 2017.
[117] Matt McFarland. Why driving slowly and responsibly can actually be fun. Washington Post. Retrieved from https://www.washingtonpost.com/news/innovations/wp/2014/06/20/why-driving-slowlycand-responsibly-can-actually-be-fun/? on 2017-01-17.
[118] Jay Goldman. Cargo2go'S Ecoscore And The Gamification. Retrieved from https://jaygoldman.com/car2go-s-ecoscore-and-the-gamification-of-driving-d6415ea1045e on 2017-01-17.
[119] Paul Barratt. Healthy competition: A qualitative study investigating persuasive technologies and the gamification of cycling. Health & place, 46:328–336, 2017.
[120] Heikki Liimatainen. Utilization of fuel consumption data in an ecodriving incentive system for heavy-duty vehicle drivers. Transactions on intelligent transportation systems, 12(4):1087–1095. IEEE, 2011.
[121] Ryosuke Ando, Yasuhide Nishihori, and Daisuke Ochi. Development of a system to promote eco-driving and safe-driving. In Smart spaces and next generation wired/wireless networking, pages 207–218. Springer, 2010.
[122] Stefan Poslad, Mohamed Hamdi, and Habtamu Abie. Adaptive security and privacy management for the internet of things (ASPI 2013). In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication, pages 373–378. ACM, 2013.
[123] Lars Holm Christiansen, Nikolaj Yde Frederiksen, Brit Susan Jensen, et al. Don't look at me, i'm talking to you: investigating input and output modalities for in-vehicle systems. InIFIP Conference on Human-Computer Interaction, pages 675–691. Springer, 2011.
[124] Arne Nykänen, André Lundkvist, Stefan Lindberg, et al. Safe and sound drive: Sound based gamification of user interfaces in cars. In Audio Engineering Society Conference: 61st International Conference: Audio for Games. Audio Engineering Society, 2016.
[125] Lotfi Asker Zadeh. Fuzzy Sets and Applications. Selected papers, ed R.R Yarger et al (John Wiley, New York, 1987).
[126] Ying Bai and Dali Wang. Fundamentals of fuzzy logic control-fuzzy sets, fuzzy rules and defuzzifications. In Advanced Fuzzy Logic Technologies in Industrial Applications, pages 17–36. Springer, 2006.
[127] Henry Nasution, Hishamuddin Jamaluddin, and Jamaluddin Mohd Syeriff. Energy analysis for air conditioning system using fuzzy logic controller. TELKOMNIKA (Telecommunication Computing Electronics and Control), 9(1):139–150, 2011.
[128] Jafar Pouramini and Ahmad Saeedi. Fuzzy model identification for intelligent control of a vehicle speed limit. Journal of Mathematics and Computer Science, 2:337–347, 2011.
[129] Fabio Tango and Marco Botta. Real-time detection system of driver distraction using machine learning. Transactions on Intelligent Transportation Systems, 14(2):894–905. IEEE, 2013.
[130] Victor Corcoba Magaña and Mario Muñoz Organero. The impact of using gamification on the eco-driving learning. In Ambient Intelligence-Software and Applications, pages 45–52. Springer, 2014.
[131] German Castignani, Thierry Derrmann, Raphael Frank, et al. Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. Intelligent Transportation Systems Magazine, 7(1):91–102. IEEE, 2015.
[132] Wenshuo Wang, Junqiang Xi, and Xiaohan Li. Statistical pattern recognition for driving styles based on Bayesian probability and kernel density estimation. arXiv preprint arXiv:1606.01284, 2016.
[133] Ahmad Aljaafreh, Nabeel Alshabatat, and Munaf S Najim Al-Din. Driving style recognition using fuzzy logic. In Vehicular Electronics and Safety (ICVES), International Conference on, pages 460–463. IEEE, 2012.
[134] Ashutosh Kumar Choudhary and Piyush K Ingole. Smart phone based approach to monitor driving behavior and sharing of statistic. In Communication Systems and Network Technologies (CSNT), Fourth InternationalConference on, pages 279–282. IEEE, 2014.
[135] Dominik Dorr, David Grabengiesser, and Frank Gauterin. Optimization of system parameters for an online driving style recognition. In Intelligent Transportation Systems (ITSC), 17th International Conference on, pages 1021–1026. IEEE, 2014.
[136] Leo Breiman, Jerome H Friedman, Richard A Olshen, et al. Classification and Regression Trees (Monterey, CA: Wadsworth and Brooks/Cole). Links, 1984.
[137] Leo Breiman. Random forests. Machine learning, 45(1):5–32, 2001.
[138] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomForest. R news, 2(3):18–22, 2002.
[139] Manuel Herrera, Luís Torgo, Joaquín Izquierdo, et al. Predictive models for forecasting hourly urban water demand. Journal of hydrology, 387(1-2):141–150, 2010.
[140] Lior Rokach. Ensemble-based classifiers. Artificial Intelligence Review, 33(1-2):1–39, 2010.
[141] Madison Schott. Random Forest Algorithm for Machine Learning. Retrieved from https://medium.com/capital-one-tech/random-forest-algorithm-for-machine-learning-c4b2c8cc9feb on 2019-04-12.

[142] Junfei Chen, Ming Li, and Weiguang Wang. Statistical uncertainty estimation using random forests and its application to drought forecast. Mathematical Problems in Engineering, 2012.
[143] Henrik Almér. Machine learning and statistical analysis in fuel consumption prediction for heavy vehicles, 2015.
[144] Jonas Lindberg. Fuel consumption prediction for heavy vehicles using machine learning on log data. PhD thesis, Ph. D. dissertation, Master's thesis, KTH, School of Computer Science and . . . , 2014.
[145] Sandareka Wickramanayake and HMN Dilum Bandara. Fuel consumption prediction of fleet vehicles using Machine Learning: A comparative study. In Moratuwa Engineering Research Conference (MERCon), pages 90–95. IEEE, 2016.
[146] Federico Perrotta, Tony Parry, and Luis C Neves. Application of machine learning for fuel consumption modelling of trucks. In International Conference on Big Data (Big Data), pages 3810–3815. IEEE, 2017.
[147] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. Statistics and Computing, 14(3):199–222, 2004.
[148] V Vapnik and Vlamimir Vapnik. Statistical learning theory Wiley. New York, pages 156–160, 1998.
[149] HyunJung Shin and Sungzoon Cho. Response modeling with support vector machines. Expert Systems with applications, 30 (4), 746–760 (2006).
[150] Steve R Gunn et al. Support vector machines for classification and regression. ISIS technical report, 14(1):5–16, 1998.
[151] Lars Rosenbaum, Alexander Dörr, Matthias R Bauer, et al. Inferring multi-target QSAR models with taxonomy-based multi-task learning. Journal of cheminformatics, 5(1):33, 2013.
[152] Dimitrios Karonis, Evripidis Lois, F anourios E Zannikos, et al. A neural network approach for the correlation of exhaust emissions from a diesel engine with diesel fuel properties. Energy & fuels, 17(5):1259–1265, 2003.
[153] Jack V Tu. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of clinical epidemiology, 49(11):1225–1231, 1996.
[154] Chihang Zhao, Yongsheng Gao, Jie He, et al. Recognition of driving postures by multiwavelet transform and multilayer perceptron classifier. Engineering Applications of Artificial Intelligence, 25(8):1677–1686, 2012.
[155] GPS Accuracy. Retrieved from https://www.gps.gov/systems/gps/performance/accuracy/ on 2018-01-09.
[156] GPS Dilution of Precision. Retrieved from https://www.e-education.psu.edu/natureofgeoinfo/c5_p21.html on 2018-02-02.
[157] Francisco Posada and John German. Measuring in-use fuel economy in Europe and the US: Summary of pilot studies. Technical report, Working Paper 2013-05). Washington, DC: International Council on Clean …, 2013. Retrieved from https://www.theicct.org/sites/default/files/publications/ICCT_FuelEcon_pilotstudies_20131125.pdf on 2018-12-11.
[158] Tsukasa Azuma. Warning! Pay Attention To These Bad MAF Sensor Symptoms (2018). Car From Japan. Retrieved from https://carfromjapan.com/article/car-maintenance/bad-maf-sensor-symptoms/ on 2019-10-29.
[159] How to Detect and Clean a Faulty Mass Airflow Sensor (2012). CarsDirect. Retrieved from https://www.carsdirect.com/car-repair/how-to-tell-if-you-have-a-faulty-mass-airflow-sensor on 2019-10-29.
[160] E. G. Laukonen, K. M. Passino, V. Krishnaswami, et al, "Fault detection and isolation for an experimental internal combustion engine via fuzzy identification," IEEE Transactions on Control Systems Technology, 3(3), 347–355 (1995).
[161] Asif Faiz, Christopher S Weaver, and Michael P Walsh. Air pollution from motor vehicles: standards and technologies for controlling emissions. The World Bank, 1996. Retrieved from http://siteresources.worldbank.org/INTURBANTRANSPORT/Resources/b02airpolution.pdf on 2018-12-11.
[162] 52north website. Retrieved from https://wiki.52north.org/Projects/EnviroCar on 2019-11-12.
[163] Diesel Engine Fuel Consumption Calculation. Retrieved from https://wiki.52north.org/Projects/DieselConsumptionCalculation on 2018-07-16.
[164] Melinda Z Acutt. Modelling greenhouse gas emissions from cars in Great Britain. Transportation Planning and Technology, 19(3-4):191–206, 1996.
[165] Google Maps Platform. Retrieved from https://cloud.google.com/maps-platform/ on 2017-02-10.
[166] Reverse geocode in R with ggmap package. Retrieved from https://www.rdocumentation.org/packages/ggmap/versions/2.6.1/topics/revgeocode on 2017-04-07.
[167] Google Maps reverse geocoding. Retrieved from https://developers.google.com/maps/documentation/geocoding/start on 2017-04-10.
[168] Darren Beams, The Need for Speed!, 2015. Retrieved from https://www.geotab.com/blog/the-need-for-speed/ on 2018-01-23.
[169] Joanine Will. What's more accurate: the car's speedo or the GPS? 2010. Retrieved from https://www.theglobeandmail.com/globe-drive/culture/commuting/whats-more-accurate-the-cars-speedo-or-the-gps/article4348596/ on 2019-11-14.
[170] Matthes Rieke, Christoph Stasch, and Christian Autermann. A Citizen Science Platform for Analyzing and Mapping de Wall, Albert Remke, Herwig Wulffius, and Simon Jirka. Citizen science for traffic planning: A practical example. In EGU General Assembly Conference Abstracts, volume 19, page 8662, 2017.
[171] Carsten Keßler and Carson JQ Farmer. Querying and integrating spatial–temporal information on the web of data via time geography. Web Semantics: Science, Services and Agents on the World Wide Web, 35:25–34, 2015.
[172] Francesco Bellotti, Sven Kopetzki, Riccardo Berta, et al. TEAM applications for Collaborative Road Mobility. Transactions on Industrial Informatics. IEEE, 2018.

[173] The International international situation for promoting eco-driving. Taiwan, 2011.
[174] Bart Beusen, Steven Broekx, Tobias Denys, et al. Using on-board logging devices to study the longer-term impact of an eco-driving course. Transportation research part D: transport and environment, 14(7):514–520, 2009.
[175] Avner Pereira, Marcel Alves, and Hendrik Macedo. Vehicle driving analysis in regard to fuel consumption using Fuzzy Logic and OBD-II devices. In Telematics and Information Systems (EATIS), 8th Euro AmericanConference on, pages 1–4. IEEE, 2016.
[176] Ebrahim H Mamdani and Sedrak Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. International journal of man-machine studies, 7(1):1–13, 1975.
[177] Craig Knott, Luke Hovell, Nathan Karimian, et al. Package 'FuzzyToolkitUoN'. University of Nottingham IMA Group, 2013.
[178] John B Heywood. Combustion in compression-ignition engines. Internal combustion engine fundamentals,, pages 522–562, 1988.
[179] He Ma. Control oriented engine modelling and engine multi-objective optimal feedback control. PhD thesis, University of Birmingham, 2013.
[180] RandomForestRegressor package. Retrieved from, http://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html on 2018-11-02.
[181] RandomizedSearchCV method of sklearn.model_selection. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html on 2018-11-10.
[182] GridSearchCV method of sklearn.model_selection. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html on 2018-11-10.
[183] Leo Breiman. Random forests Leo Breiman and adele cutler. Random Forests-Classification Description, 2015.
[184] Terence Parr, Kerem Turgutlu, Christopher Csiszar, et al, "Beware Default Random Forest Importances," 2018 Retrieved from https://explained.ai/rf-importance/index.html on 2018-11-12.
[185] Sujata Khedkar, Akshaykumar Oswal, Manjari Setty, et al. Driver evaluation system using mobile phone and OBD-II system. Int. J.Comput. Sci. Inf. Technol, 6(3):2738–2745, 2015.
[186] Yi Lu Murphey, Robert Milton, and Leonidas Kiliaris. Driver's style classification using jerk analysis. In Workshop on Computational Intelligence in Vehicles and Vehicular Systems, pages 23–28. IEEE, 2009.
[187] Adrian B Ellison, Stephen P Greaves, and Michiel CJ Bliemer. Driver behaviour profiles for road safety analysis. Accident Analysis & Prevention, 76:118–132, 2015.
[188] John Jairo Martinez Molina. Commande de l'inter-distance entre deux véhicules. PhD thesis, 2005.
[189] SVR package of scikit-learn. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html on 2018-12-23.
[190] Weiliang Zeng, Tomio Miwa, and Takayuki Morikawa. Exploring trip fuel consumption by machine learning from GPS and CAN bus data. Journal of the Eastern Asia Society for Transportation Studies, 11:906–921, 2015.
[191] Keras: The Python Deep Learning library. Retrieved from https://keras.io/ on 2019-10-10.