Vehicular Data Space: The Data Point of View

Paulo H. Rettore, Guilherme Maia, Leandro A. Villas, and Antonio A. F. Loureiro

***Abstract*—Over the years, governments and automakers launched initiatives to improve road traffic efficiency, safety, and people mobility. They have been working on various aspects of intelligent transportation systems (ITSs), which aim to improve decision-making, availability of information, and communication technologies to provide applications and services to boost the transportation systems. The development of new applications and services for ITS depends on the availability of different data sources, what it is not the current case. Many studies focus on the communication issues of applications and their associated challenges. To reveal the recent vehicular data use, we examined the most remarkable studies of the last few years, which describe services and applications for ITS, however with a focus on the data employed by them. We introduce the concept of vehicular data space (VDS), which is then used to describe the vehicular scenario from the perspective of data. Moreover, we outline a taxonomy, according to the different data sources. We also categorize the applications, highlighting the data each one used in their approach. Finally, we present some challenges and open issues related to the process for *data creation*, *data preparation*, *data processing*, and *data use*. In a nutshell, this paper constitutes one of the first holistic surveys on services and applications for ITSs, focusing on the data used by them, as well as their future challenges.**

***Index Terms*—Vehicular data space (VDS), vehicular sensor data, connected vehicles, intelligent transportation system (ITS), heterogeneous data, vehicular ad-hoc network (VANET).**

I. Introduction

MORE and more cities will have to deal with significant issues related to transportation and traffic because people and goods are in constant need for quicker and safer mobility modes. The number of fatalities and injuries on the road has achieved an alarming scenario. Globally, 1.3 million people die every year, and up to 50 million suffer severe injuries. These facts have a direct impact on the economics of nations, leading to costs in the order of about 2% to 5% of the Gross Domestic Product (GDP) in many countries [1]. It is also reported that traffic congestion results in critical economic and environmental costs. In 2011, 498 U.S. urban areas were evaluated regarding the congestion impact. It was found that about USD 121 billion was wasted on fuel consumption and more than 25 billion kg of CO_2 was emitted. Those values were USD 24 billion and 4,53 billion in 1982, respectively. In 2014, 471 U.S. urban areas were observed, and the costs related to wasted fuel consumption due to congestion reached USD 160 billion [2], [3].

Over the years, governments and car manufacturers launched initiatives to improve road traffic efficiency, safety, and mobility. They have been working on various aspects of Intelligent Transportation Systems (ITSs), which aim to improve decision-making, availability of information and communication technologies to provide applications and services to boost the transportation systems. Some initiatives are described in [4], [5], [6], [7], [8], [9], [10], [11], [12]. There is clearly a considerable growth of on-board informatics inside vehicles. Currently, each vehicle has an average of 60-100 embedded sensors, and these numbers can go up to as much as 200 sensors per vehicle in 2020. Moreover, according to Machina Research [13], in 2020, about 90% of new cars will feature an Internet-integrated, while it was about 10% in 2013.

Given the importance of sensors to a vehicle's operation, new models embed many high-quality sensors [14] to get more reliable and diverse information about themselves. In that way, Advanced Driver Assistant Systems (ADAS) offer a means to enhance, among other things, the driver's safety and comfort [15]. In the last years, the development of vehicular sensors had a significant increase. As a consequence, the number of connecting cables inside the vehicle has also increased, resulting in an additional 50 kg to the vehicle mass, besides the increase of the final vehicle cost, and the difficulty of installing and maintaining all systems working properly [16]. For that reason, an Intra-Vehicle Sensor Network (IVSN)[1] may need to rely on wireless communication for its operation. Thus, the Intra-Vehicle Wireless Sensor Network (IVWSN) is a research topic in the field of vehicular sensor communication.

An important issue here is how to have a wireless connection among sensors and the Engine Control Unit (ECU). This sensor network usually has some particular characteristics, such as sensors are stationary and are only one hop away to the ECU, and have no energy constraint. In spite of these characteristics, there are some challenges related to the efficient use of wireless channels, such as latency, reliability, security and interference issues in a dense urban scenario. In particular, we are interested in challenges and opportunities related to the whole data space that influences or is influenced by vehicles. How those sensors communicate, wired or wireless,

Manuscript received July 25, 2018; revised January 26, 2019; accepted March 25, 2019. Date of publication April 22, 2019; date of current version August 20, 2019. This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), in part by the Conselho Nacional de Desenvolvimento Científfico e Tecnológico (CNPq), in part by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), in part under Grant #2019/06056-4, Grant #15/24536-2, and Grant #15/24494-8, and in part by São Paulo Research Foundation (FAPESP).
(Corresponding author: Paulo H. Rettore.)

P. H. Rettore, G. Maia, and A. A. F. Loureiro are with the Computer Science Department, Federal University of Minas Gerais, Belo Horizonte 31270, Brazil (e-mail: rettore@dcc.ufmg.br; jgmm@dcc.ufmg.br; loureiro@dcc.ufmg.br).

L. A. Villas is with the Institute of Computing, University of Campinas, Campinas 13083-970, Brazil (e-mail: leandro@ic.unicamp.br).
Digital Object Identifier 10.1109/COMST.2019.2911906

[1]Also mentioned as Intra-Vehicular Communication (IVC) and Intra-Vehicular Network (IVN).

to provide useful data is not the focus of our study. For more details, see [17], [18], [19], [20], [21], [22], [23], and [24] for a broad comprehension of Intra-Vehicle Networks.

The development of new applications and services for ITS depends on the availability of different data sources, what it is not the current case. In fact, many data sources may play a central role in the development of new solutions, tools and businesses. In the literature, there are some studies describing the main features and properties of ITS applications [16], [25], [26]. In this work, we survey recent proposals describing services and applications for ITS, but with a focus on the data employed by them. We introduce the concept of Vehicular Data Space (VDS), which is then used to describe the vehicular scenario from the perspective of data. Moreover, we outline a taxonomy and applications based on that concept, and we end with the challenges and open issues based on the data cycle on the VDS.

We conducted a literature review in the area of ITS scope, focused on taxonomies of data and applications, aiming to answer the question: *"What are the data sources which feed the ITSs and how they can be used to provide applications and services to ITS?"*. As a result, we highlight, in summary, the following contributions: (i) The data viewpoint of ITS' applications, introduced by an original concept, Vehicular Data Space (VDS). We draw the attention of researchers to the importance of data and their availability and quality to ITSs. We reinforce the idea that without due care of data, the communication aspects are useless; (ii) The identification of the set of data currently used by ITS' applications, including those ones less explored, and the importance of heterogeneous data fusion to the design of more robust and efficient classes of applications; (iii) The relevance in understanding the use of data in ITSs.

The rest of the paper is organized as follows. In Section II, we introduce the VDS concept and discuss the methodology used to identify relevant studies in the literature. In Section III, we present an example of the VDS environment and its respective entities and data. In Section IV, we present a taxonomy of the vehicular data space from the perspective of data sources, and analyze existing solutions. In Section V, we discuss some potential applications in VDS, focusing on the data point of view. In Section VI, we examine the challenges and open issues in the use of vehicular data, based on the data cycle of the VDS. Finally, in Section VII, we conclude the survey with some possible future directions.

II. Vehicular Data Space

Given the importance of data to ITS, this work looks at the ITS field using the perspective of data. For that, we categorize existing literature research according to the data sources employed by them. The aim here is to consider different data aspects, such as availability, spatiotemporal correlations, acquisition challenges, frequent used data types and their applicability, and heterogeneous data fusion issues. Therefore, our goal is to present the vast ITS field according to the vehicular data context.

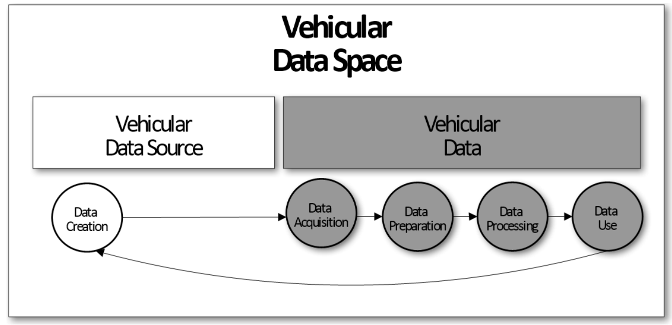


Fig. 1. The big picture of Vehicular Data Space and its respective state of data cycle in the VDS.

For that, we introduce the concept of a VDS, which covers the various aspects regarding data to provide a descriptive view of the transportation scenario, however, differently from the approach presented in [16]. Here, we assume that a VDS encompasses both the data sources and the data produced by them. Hence, we conduct a literature review focusing on the concepts of Vehicular Data Source (VDSource), Section IV, and Vehicular Data (VD), Section V. Besides, we created the data cycle for VDS, aiming to show stages which may serve as a guideline to propose new solutions to the ITS scenario and allow a whole comprehension of the VDS. Figure 1 summarizes the subsets of VDS and the five stages of the proposed data cycle, which span from the data creation to their use. It is important to notice that the focus of this work is not to outline each stage of the data cycle, but to show its role in the VDS to later guide us in the discussion of challenges and open issues. Each subset in the VDS can be briefly described as follows:

- Vehicular Data Source (VDSource)
  - *Data Creation:* The process of sensing environment variables through real or virtual sensors.
- Vehicular Data (VD)
  - *Data Acquisition:* The process of making these data available through device logs, storage, cloud or even APIs.
  - *Data Preparation:* The filters or corrections applied to the data so it can be processed.
  - *Data Processing:* The methods and algorithms applied to the data according to its properties and desired use.
  - *Data Use:* The proposed use (e.g., applications) which may power other data cycles or applications.

Based on that, the VDSource deals with the *Data Creation*, whereas the VD covers the rest of the data cycle, i.e., *Data Acquisition*, *Data Preparation*, *Data Processing* and *Data Use*, allowing the developing of services and applications for ITSs. As mentioned, it is out of our scope of this work to provide a deep discussion about each of these steps in Section V (application section), except the *Data Use*, which discusses how the data may be used, disregarding its acquisition and processing. Furthermore, we discuss the challenges and open issues (Section VI) considering the data cycle of VDS in mind.

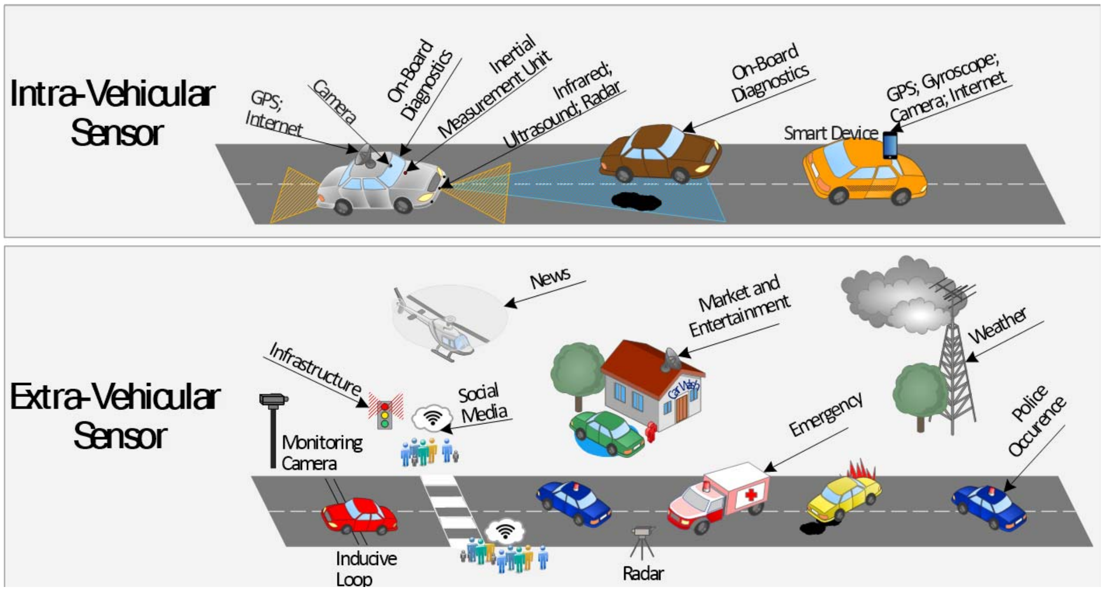


Fig. 2. Vehicular data space provided in the urban area.

III. Entities of the Vehicular Data Space

In this section, we present an example of the VDS environment and its respective entities and data.

A. Overview of VDS

Vehicular *Ad-hoc* Networks (VANETs) are a derivation of Mobile *Ad-hoc* Networks (MANETs), in which vehicles are equipped with computing, sensing and communication capabilities [27], [28], [29]. Moreover, VANETs possess characteristics that are specific to the vehicular environment, such as vehicles are expected to move in well-defined patterns and concentrate in high-density urban regions, and vehicles have a more predictable mobility model. Built on top of VANETs, the Vehicular Sensor Network (VSN) [30], [31] is a powerful sensing platform that provides the capability for collecting, computing and sharing sensor data. A vehicle contains various types of highly reliable sensors and almost eliminates the energy constraints of traditional MANETs, due to its rechargeable battery. Moreover, vehicles can leverage the communication capabilities already deployed in urban areas, such as cellular and wireless networks.

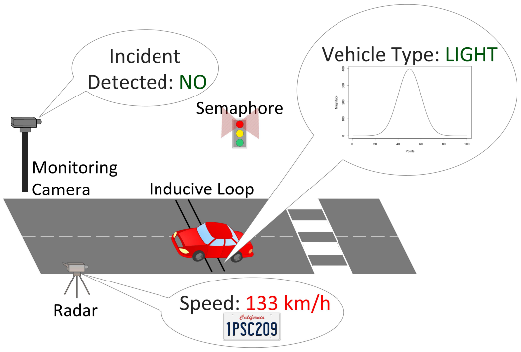
The perception of the surrounding environment is paramount for provisioning many services in VANETs. Physical sensors play an important role in control systems, as they provide data on operational states and malfunctions of monitored entities. Vehicular control systems are among those that depend on sensor data to actuate on their components to provide a safe and enjoyable driving experience. Traffic control systems also depend on sensor data to measure the vehicle flow, traffic lights coordination, and delays. Weather monitoring systems rely on sensors for predicting storms. Moreover, Participatory Sensor Networks (PSN) also play a relevant role in monitoring and control systems in a wide scope. News and Social Media can act as a virtual sensor wherever there is a lack of physical sensors. For instance, an accident report can be filled out by Social Media users in areas with no road sensors infrastructure. Moreover, people's feelings who pass near an incident cannot be perceived by physical sensors.

Many studies in VANETs focus on the communication issues for ITS and their associated challenges. For instance, assume an accident between two vehicles. Most studies are interested in knowing how this event can be disseminated through a road to alert other drivers and the road administrators, i.e., how to efficiently broadcast the emergency event. On the other hand, here we focus on the data. In other words, both vehicles are constantly producing data. Therefore, how can such data be used to improve an accident avoidance system? Furthermore, how can the road historical data be analyzed to reduce the risks of an accident?

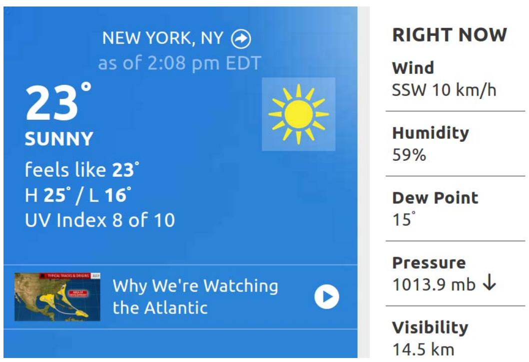
We consider as a VDS all data used to provide a descriptive view of a vehicular scenario, such as intra-vehicle data, traffic flow data, traffic incidents data, infotainment and others. Notice that the data may be produced by intra-vehicle sensors, smart devices or even social media, for instance. The first step before proposing solutions for ITS is to understand the data and its sources, such as the entities responsible for acquiring and, in some cases, providing data access to the community. We show an example of the VDS and its respective entities, which produce data in an urban area, in Figure 2. Figure 2 shows an example of the VDS and its respective entities, which produce data in an urban area. In the following, we describe some of data sources shown in this figure, grouping them in Infrastructure, Transit Authority, Vehicle, Publicity and Media. We highlight that the concept of data in our context may be related to the raw data or also data in a given context, i.e., a piece of information.

B. Infrastructure

Infrastructure data address a range of sensors, such as vehicle detection loops, called inductive loop traffic detectors,

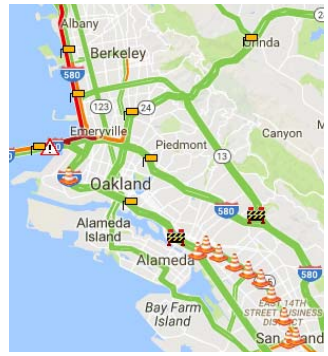


(a)

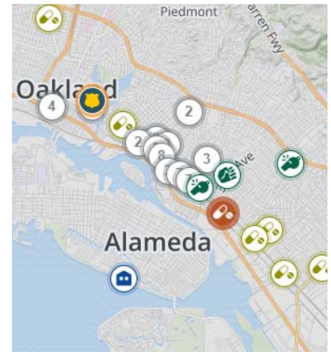


(b)

Fig. 3. Data provided by infrastructure. (a) Data provided by the inductive loop, monitoring cameras and radar. (b) Data provided by a weather station in New York City [33].



(a)



(b)

Fig. 4. Data provided by government entities. (a) Road conditions, incidents and traffic level provided by the U.S. Department of Transportation. (b) Data provided by the government and available through a Crime Reports Platform [34].

monitoring cameras, radars, traffic lights, and weather sensors. Inductive loops are based on a wired electromagnetic communication (see the black lines on the roads in Figure 2). They are installed on the pavement and can detect a vehicle passing at a certain point and its speed. Inductive loops have also been used to classify types of vehicles, based on their signatures [32].

Similarly, however, with a higher deployment cost, monitoring cameras or radars can also be used to detect the speed of a vehicle or its type. Moreover, cameras have also been used to detect and prevent accidents, and to broadcast notifications to authorities. A preventive situation can be illustrated by an animal that crosses a road, and, then, the authorities are promptly notified about it, so they can take actions to avoid future accidents. Cameras can also record the vehicle's license plate when traffic rules are broken (e.g., the red car crossing a red traffic light in Figure 2). The combination of inductive loops, traffic lights, cameras and radars produces a virtual sensor that allows traffic agencies to apply the governing legislation and eventually issuing traffic tickets. Figure 3a shows each data source just mentioned.

The road infrastructure needs to work together to prevent traffic jams and high traffic flow. For instance, a traffic light can be based on static time intervals, or adapt its behavior according to the perceived traffic conditions. The data traffic may be collected as a result of wired or wireless communication with other traffic lights, inductive loops and radars. Other data provided by the infrastructure are weather stations, which provide, in real time and for a certain area, data about temperature, pressure, wind speed, dew point, humidity, and also prediction data on the chances of precipitation. Figure 3b shows an example of a New York weather station[2] in Figure 3b.

C. Transit Authority

Government entities play an essential role in the transportation system since they help decision makers and overall people to better understand the mobility behavior in a city. Most countries possess agencies that provide traffic-related data, such as statistics about traffic jams, accidents, road state, police occurrences, medical occurrences, fatalities, and injuries on the road, and mobility patterns. Such data may be used by different stakeholders to make informed decisions. For instance, in possession of data about fatalities and injuries on a specific road, drivers can change their actions and drive more carefully. As an example of government data, Figure 4a shows the traffic alerts provided by the U.S. Department of Transportation (DoT) in the state of California, aiming to show blocked roads, incidents, traffic intensity and alerts to road users.

Figure 4b shows another type of data provides by Police Departments. Using an online platform, Socrata [34] makes government data available to citizens. Crime Reports show a variety of crimes, such as disorder, vehicle thefts, property crime, robbery, sexual offense and drugs. Such data allow users to better understand a particular area. Notice that the data provided by these entities may not be the raw data. Some treatment may be introduced to offer a more detailed scenario. Despite this, we still consider them as data.

D. Vehicle

An important data source in a VANET scenario is the vehicle itself. Vehicles have sensors to collect data about speed, acceleration, movement, luminosity, location, the presence of people or obstacles, external and internal temperatures, and

[2]https://weather.com

TABLE I
DATA FROM A VEHICLE AND ADDITIONAL DEVICES EMBEDDED IN IT

Vehicular Sensor Data									
From Additional Devices				*From Engine Control Unit*					
Time	Obstacles Detection	Video Record	Road Condition	Throttle Position	Tire Pressure	Fuel Level	Torque	Engine RPM	Acceleration
Location	3-axis Acceleration	Audio Record	Atmospheric Pressure	Steering Wheel Angle	Battery Voltage	Intake Air Temp	Fuel Flow	Speed	Light
GPS Speed	Altitude	Ambient Air Temp		Fuel Consumption	Gear	Trip Distance	Engine Coolant Temp	CO_2	Air Conditioner Temp

current structural state, which can provide information to alert the driver about events about the vehicle. Moreover, sensors may be used to control the operation of vehicles. For instance, data provided by the luminosity sensors can control the automatic functioning of the lights, turning them on during the night. Furthermore, proximity sensors can help drivers to keep a safe distance from neighboring vehicles, avoiding collisions. These sensors play an important role in autonomous vehicles. Table I presents some data that can be acquired directly from the ECU of vehicles or additional devices embedded in vehicles.

Sensors embedded in a vehicle can also be used to detect many events in the surrounding environment during the vehicle's trajectories. Using the On-Board Diagnostic (OBD), data collected from sensors can be used to monitor the traffic and events around the city. For instance, the vehicle's GPS data can support a traffic monitoring service, alerting about traffic jams. In another scenario, combining data from both accelerometer and GPS, it is possible to monitor the presence of holes on the roads.

E. Publicity

The VDS also contains data provided by market and entertainment companies. These data aim to offer personalized products, services or comfort applications to the drivers. Figure 2 shows a simple example of a market on the road, where a Car Wash company tries to sell its services to vehicles that will pass in front of its location, using a Vehicle-to-Infrastructure (V2I) infrastructure. Based on that same idea, a car maintenance company can offer services to the driver since the vehicle sends data about its state and eventual malfunction to the car manufacturer.

A variety of applications can be developed to provide entertainment to the passengers of a vehicle, based on information about them and their vehicles. For instance, their smartphones carry a personal user data and applications which become useful through the dashboard display and multimedia kit inside the cars. This allows a better involvement between passengers and the environment around them. There are private companies with initiatives, focusing on connecting costumers with their cars, growing the comfort and the client satisfaction. For instance, the General Motors developed OnSart,[3] Audi offers Audi Connect,[4] Apple developed CarPlay,[5] Google developed Android Auto,[6] and Toyota and BMW have also an infrastructure for their users, Toyota Touch 2[7] and BMW ConnectedDrive,[8] respectively.

F. Media

The growth and popularity of the Internet implied the increase of media in reporting the conditions of transportation. The incidents, traffic conditions, the number of fatalities, road conditions, the events in a given location and so on become the goal of many types of media, i.e., social media, news blogs, newspapers, map navigation and transit insights, radios, and TVs. Constituting, a relevant way to disseminate and provide information to the better comprehension of the transportation system. Even though the data provided by media can be subjective and biased, those data can provide information difficult to obtain with other data sources.

The use of social media is a novel possibility to obtain information about the traffic and road conditions, or report events to other drives. These are particular Location-Based Social Media (LBSM) apps, which enable mobile users to act as mobile sensors, monitoring the environment, weather, urban mobility and traffic conditions. The main feature of this data type is the real-time information of the sensed events. Typically, users retrieve the accurate data about the traffic conditions. Another important feature is its large coverage, since all users connected to the network can access these data with no restrictions. Figure 5 shows examples of types of media used in the VDS in benefit of applications to the ITSs. Figure 5a shows textual data provided by reports of the user from the Twitter Platform,[9] whereas Figure 5b displays visual data provided by a combination of users' reports of the Waze[10] app, allowing other users to have a better overview of the traffic conditions.

A different way to obtain data of VDS comes from radio stations created to disseminate information about the road state. For instance, there are radio stations focused on broadcasting

[3]https://www.onstar.com/us/en/home.html

[4]https://www.audiusa.com/help/audi-connect

[5]https://www.apple.com/ios/carplay/

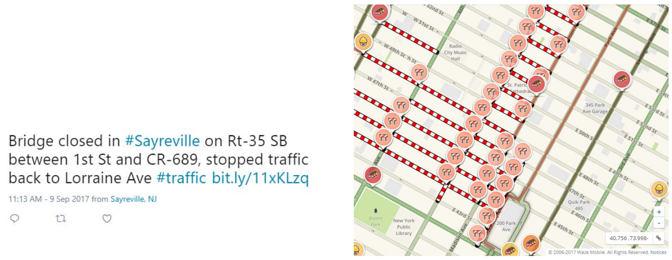
[6]https://www.android.com/auto/

[7]https://www.toyota-europe.com/world-of-toyota/articles-news-events/2016/toyota-touch-2

[8]http://www.bmwusa.com/standard/content/innovations/bmwconnecteddrive/connecteddrive.aspx

[9]https://twitter.com

[10]https://www.waze.com/



(a) (b)

Fig. 5. Data provided by media. (a) Data provided by the LBSM Twitter [35], reporting the traffic occurrence in NY City. (b) Data provided by Waze map [36] in NY City.

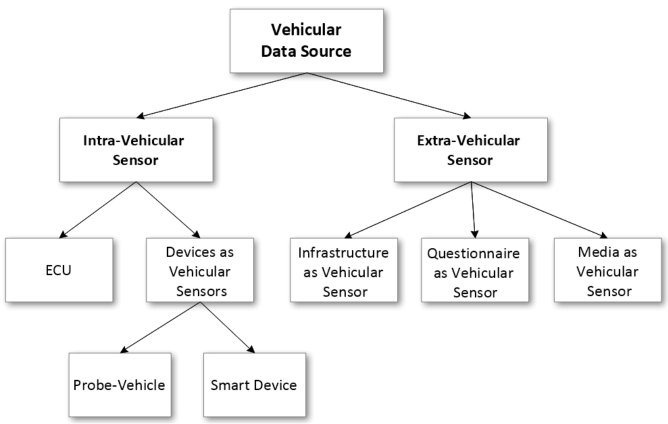


Fig. 6. Taxonomy of vehicular data space based on the point of view of the source.

information about the road conditions like a road blockade, accident and animals on the road. These pieces of information are obtained from drivers' notifications and road employee observations.

IV. Taxonomy of Vehicular Data Source

In this section, we categorize the Vehicular Data Source (VDSource) into two main groups named Intra-Vehicular Sensors (Section IV-A) and Extra-Vehicular Sensors (Section IV-B), as shown in Figure 6. Afterwards, we discuss each leaf of the taxonomy tree and present an overview.

A. Intra-Vehicular Sensor

Intra-Vehicular Sensor (IVS) corresponds to the subset of sensors that describe the main interactions between a vehicle and its driver, passengers or its surrounding environment, from the perspective of the vehicle itself. In other words, IVS represents all sensors embedded in a vehicle or on-board that measure the vehicle state, the drivers' behavior or the environment conditions in its surrounding.

IVS may collect data from the ECU, such as engine load, engine coolant temperature, engine Revolutions Per Minute (RPM), vehicle speed, throttle position, and others. Moreover, IVS may also collect data provided by devices on-board of a vehicle. These devices are classified as Device as Vehicular Sensor (DVS). We further categorize these devices into *Probe-Vehicle*, where a set of precise sensors are used to monitor a particular event, and *Smart Device*, where devices, such as smartphones, tablets and other pieces of hardware act as data sources. In the following, we group the proposals according to the type of IVS they employ to collect data.

1) Engine Control Unit: Given the importance of sensors to a vehicle's operation, new models embed many high-quality sensors to get more reliable and diverse information about themselves. All data produced by sensors in a vehicle are delivered to its ECU through an internal network, named Controlled Area Network (CAN), which is accessible through the vehicle's OBD port. A useful analogy is to suppose that the OBD is the language that we use to speak about a vehicle's state, as informed by the ECU, using a communication device (CAN).

The OBD system was first introduced to regulate emissions. However, it is now used for a variety of applications. There are different signaling protocols to transmit internal sensor data to external devices through a universal port. Such a universal port is present in all cars produced since 1996 in the U.S. and Europe. There are Parameter IDs (PIDs) to access sensor information using the OBD, which identify individual sensors. Some PIDs are defined by regulatory entities and are publicly accessible. However, manufacturers may include other sensors' data under specific and undisclosed PIDs.

The 52 North Initiative for Geospatial Open Source Software [37] proposed a platform named EnviroCar for collecting geographic data and vehicles' sensors. The EnviroCar is an open platform for Citizen Science projects, which aims to provide sustainable mobility, traffic planning and share the findings from the industry when collecting and analyzing car data. Using an OBD adapter into a car, they collected a variety of sensor data and uploaded it to the Web. The system consists of the EnviroCar app and the EnviroCar server. Bröring *et al.* [37] described the spatiotemporal RESTful Web Service interface and the designed data model. Since 2015, there are over 500,000 measurement data points collected and these numbers are continuously growing. Reininger *et al.* [38] described a prototype to provide vehicular data access through a website. Using an OBD port and a smartphone, they provided data, such as speed, RPM, fuel consumption, coordinates, and altitude, for later post-processing and analysis. They also described a sandboxing mechanism that prevents malicious attacks from other programs on the smartphone.

Van Ly *et al.* [39] showed the potential of using inertial sensors to distinguish drivers. They concluded that the acceleration feature does not play a significant role in such process, contrarily to the braking and turning features. As an experimental test-bed, they employed a LISA-X (probe-vehicle) to acquire all their data. This experimental vehicle was outfitted with a variety of sensors and vision system. They used signals from a CAN, such as an engine speed, brake pressure, acceleration, pedal pressure, vehicle speed and angular rotation to recognize the vehicle maneuvers represented by three types of events: braking, acceleration, and turning. D'Agostino *et al.* [40] proposed a classification method for identifying driving events using short-scale driving patterns. For that, they relied on data provided by CAN and GPS.

Carmona *et al.* [41] proposed a novel tool to analyze the driver's behavior and identify aggressive behavior in real time. For that, they relied on a variety of data, such as brake usage frequency, throttle usage, engine RPM, speed, and steering angle. Such data were retrieved using a Raspberry Pi device connected to the CAN through an OBD port. Kumtepe *et al.* [42] developed a solution to detect the driver's aggressiveness in a vehicle using visual information and in-vehicle sensor data acquired from the CAN, such as vehicle speed and engine rotation (RPM). They could detect aggressive driving behavior with a success rate of over 93%.

Johnson and Trivedi [43] showed that sensors available on smartphones can detect movement with a similar quality to a vehicle CAN bus, allowing the recognition and recording of driver's actions. However, Paefgen *et al.* [44] showed that such quality depends on the smartphone positions and the type of event being identified. AbuAli [45] collected data from vehicular sensors using an OBD port to detect the driver's behavior, road artifacts and accidents. To address these issues, it was used the vehicle speed, throttle position, RPM and coordinates to track the vehicle's location. That work showed that the proposed system can detect road artifacts with a success rate of about 84%.

Zhang *et al.* [46] developed a driver's identification model using sensors available both on mobile phones and vehicles, in which data was collected through an OBD port. They evaluated three vehicles in two different environments, a controlled and a naturalistic. Considering only the vehicular sensors, such as acceleration pedal position D, throttle position manifold, absolute throttle position B, relative throttle position, acceleration pedal position E, engine RPM and torque, the classification model obtained a 30.36% accuracy in the controlled environment with 14 drivers whereas in the naturalistic environment with two drivers per vehicle it obtained an 85.83% accuracy. Satzoda and Trivedi [47] proposed techniques to extract semantic information from raw data provided by vehicles in order to minimize the effort needed for data reduction in Naturalistic Data Studies (NDS). They applied fusion techniques to data from a forward-looking camera, vehicle's speed from a CAN bus, and Inertial Measurement Unit (IMU) and GPS as well. As result, they extracted a set of 23 pieces of semantic information about the location and position of the vehicle on the lane, its speed, the traffic density and the road curvature.

Magaña and Muñoz-Organero [48] proposed a solution to reduce the impact of traffic events on fuel consumption. For that, they first developed a system to detect traffic incidents based on the rolling resistance coefficient, the road slope angle and the vehicles speeds. Next, they found an optimal deceleration by anticipating traffic incidents, improving fuel consumption by up to 13.47%. Through an OBD port, they obtained the vehicle speed, acceleration, engine speed and the fuel consumption. Meseguer *et al.* [49] developed a smartphone app aiming to characterize the road type as well as the aggressiveness of each driver. For this purpose, they relied on data, such as speed, acceleration, and RPM acquired from the CAN. As result, they achieved an accuracy of 98% when attempting to characterize road types and 77% when characterizing the driving style. Similarly, Hong *et al.* [50] developed a platform to model aggressive driving styles based on data from smart devices and ECU. From a smartphone, they used GPS location and 3-axis acceleration. From the IMU, they employed the number of turns and acceleration, whereas from the vehicle they used the speed, engine RPM and throttle position. In addition, they employed the Manchester Driving Behavior Questionnaire (DBQ) to complement the characterization of the driving style. As a result, using all three data sources, their prediction achieved 90.5% accuracy, while the questionnaire data achieved 81%.

Hallac *et al.* [51] developed a method for predicting the identity of drivers based on in-vehicle sensor data collected from a CAN. In particular, they used the steering wheel angle, steering wheel velocity, vehicle speed, brake pedal position and gas pedal position. The results achieved an accuracy of about 76.9% for a two-driver classification and 50.1% for a five-driver classification. Martinez *et al.* [52] proposed a non-intrusive method for identifying impostor drivers. They relied on a dataset [53] that allowed access to a variety of sensor data. However, a reduced set of variables from the CAN was used, such as RPM, brake pedal and throttle position. As result, they achieved an identification rate greater than 80% for every evaluated group category.

Riener and Reder [54] conducted a study aiming to show that traffic safety and efficiency improve when competent drivers support the not so competent ones by sharing the road and driving data. The data acquisition was made using the OpenXC Platform [55] and a smartphone. They used the steering wheel angle, torque, RPM, vehicle speed, throttle position, fuel consumption, gear position, GPS and 3-axis acceleration. They developed a social driving app that provides recommendations about how to drive on a given track based on experiences shared by other drivers. Rettore *et al.* [56] explored the driver's identification as an extra authentication factor to local services and vehicular networks. In this respect, they developed a virtual sensor to determine the driver's identity (legitimate or suspect), with a precision above 98%, using embedded sensor data such as vehicle speed, fuel flow, gear, engine load, throttle position, emissions and RPM.

Rettore *et al.* [57] also developed a virtual gear sensor for manual transmission cars, which allows to relate each gear with the fuel consumption. They proposed a methodology to recommend the best gears according to current speed and torque. Using such methodology, they were able to reduce the fuel consumption and the CO_2 emissions by approximately 29% and 21%, respectively. They collected data from vehicle sensors, such as engine load, engine RPM, fuel flow, throttle position, trip distance and CO_2 through an OBD port. Rutty *et al.* [58] conducted a study to show the impact of eco-driving training in a municipal fleet. They used the CarChip [59] technology to acquire data from the CAN and evaluate their proposal. The results showed an average decrease of engine idling between 4% and 10%, and an average reduction of emissions of 1.7 kg of CO_2 per vehicle per day. One year later, Rutty *et al.* [60] assessed the value of vehicle monitoring technology (VMT) and eco-driver training to reduce emissions and fuel. They showed the results of eco-driving training in a fleet of vehicles at the ski resort operation in Ontario, Canada. The fleet reduced 14% of their average daily speed, 55% of abrupt deceleration, 44% of hard accelerations, and 2% of idling time. Finally, they achieved a decrease of 8% in fuel costs and CO_2 emissions.

Similarly, Ayyildiz *et al.* [61] developed an advanced telematics platform to compare the driving style before and after eco-driving training. They acquired data from an OBD port, such as vehicle speed, fuel consumption, emissions and GPS location using a smartphone. The study presented a reduction of 5.5% in fuel consumption for heavy vehicles, while light vehicles did not show significant variations. Brace *et al.* [62] proposed an onboard Driver Assistant System (DAS), which encourages to improve the driver's driving style. Specifically, the system aims to decrease fuel consumption by reducing the rates of acceleration and early gear changes. For that, they employed data from the vehicle ECU. The used data include vehicle speed, throttle position, engine speed, engine load, engine fueling demand and engine coolant temperature for a total of 39,300 km of collected trip data. They showed fuel savings of up to 12% and an average fuel savings of about 7.6%. Zhao *et al.* [63] proposed and evaluated the Dynamic Traffic Signal Timing Optimization Strategy (DTSTOS), aiming to reduce the total fuel consumption and traffic delays in a road intersection. Using the VISSIM traffic simulator [64], they obtained data, such as vehicle speed and fuel consumption.

Araújo *et al.* [65] proposed a smartphone app to help drivers to change their behavior and, consequently, reduce the fuel consumption. For that, they used the vehicle state data acquired from the CAN bus, through an OBD and the smartphone sensors. They relied on data, such as vehicle speed, acceleration, altitude, GPS, throttle position, instant fuel consumption and the engine rotations. Andrieu and Pierre [66] developed an efficient Ecological Driving Assistance System (EDAS) aiming to detect eco-driving behavior and provide drivers with recommendations to help them to reduce the fuel consumption and preserve their safety. They used the CAN and OBD to monitor driving parameters, for instance, vehicle speed, RPM, fuel, brake pedal and throttle position. They showed that it is possible to reduce fuel consumption just by following simple rules of eco-driving. After applying those rules, the average fuel consumption, the speed, and the time spent above the legal speed limit reduced approximately 12.5%, 5.8% and 30%, respectively.

Paefgen [67] conducted a study aiming to determine the risk of an accident according to collected vehicular sensory data. Focusing on the automobile insurance market and aiming to introduce adaptive insurance tariffs, known as Pay-As-You-Drive (PAYD), the author used a dataset of location trajectories and vehicle's speed data from an OBD port to develop an algorithm to reconstruct trajectories when GPS data were missing. The result was a business model for insurance telematics offerings.

2) Probe-Vehicle: A Probe-Vehicle is a vehicle specifically designed for collecting traffic data, road data, driver data and other types of data in real-time. Its main feature is the high quality of sensors embedded in it. For that reason, many public and private initiatives use that kind of vehicle to measure the quality of roads, weather and driver's behavior. In the following, we analyze studies that employed probe-vehicles to achieve their goal.

Mednis *et al.* [68] designed an embedded device (CarMote) that focus on monitoring road surface and weather. They used a microphone, accelerometer, temperature and humidity sensors to create a detailed map of the road quality and meteorology. Van Ly *et al.* [39] collected sensor data from the front side radar, front/rear camera, lateral (Left/Right) and longitudinal (Forward/Backward) acceleration and Yaw Angular Velocity sensors to describe three types of events: braking, acceleration and turning. Satzoda and Trivedi [47] associated inertial data from the IMU, GPS and camera with the vehicle speed obtained from its CAN bus. Beyond the in-vehicle data used by D'Agostino *et al.* [40], they also used a camera, aiming to record the trips and label the main events while en-route.

Guo and Fang [69] conducted a study aiming to identify features associated with dangerous driving. Using demographic, personality and driving characteristic data, they predicted who the high-risk drivers are. The authors used the first large-scale study conducted in the United States in 2006, the 100-Car Naturalistic Driving Study (NDS), to develop their methodology and application. The vehicles were instrumented with a set of sensors, such as five camera views around the vehicle, GPS, speedometer, three-dimension accelerometer, radar, and others. The data were collected continuously for 12 months with approximately 43,000 hours and 2 million vehicle miles. The results associated the driver's age, personality and critical incident rate with the risk of crashes and near-crash events. They also showed that approximately 6% of drivers are high-risk drivers, 12% are moderate-risk while 84% are low-risk.

Elhenawy *et al.* [70] introduced a new predictor for driver's aggressiveness and demonstrated that this measure enhances the modeling of driver stop/run behavior. They also developed a model that can be used by traffic signal controllers to predict the driver's stop/run decisions. The vehicles were equipped with a Differential Global Positioning System (DGPS) unit, a longitudinal accelerometer, acceleration and brake pedal position, and, in some cases, cameras as well. Carmona *et al.* [41] also used in their analysis of the driver's behavior a DGPS, which is composed of a base station that provides improved location accuracy in real-time. They also used an IMU, which has embedded accelerometers and gyroscopes.

Relying on visual information, Kumtepe *et al.* [42] developed a method to detect the driver's aggressiveness by detecting lane deviation and collision time. Andrieu and Pierre [66] employed a GPS, front car camera and a fuel flow meter to develop an efficient EDAS. They also used a specific fuel flow hardware aiming to validate the fuel consumption provided by an OBD port. In this direction, Honda Sensing [71] is an example of a practical solution currently available for their customers. Since 2015, Honda embeds in its cars a suite of safety and driver-assistive technologies such as Collision Mitigation Braking, Road Departure Mitigation, Adaptive Cruise Control, Lane Keeping Assist, Traffic Sign Recognition and Auto High-Beam Headlights.

3) Smart Device: Similarly to probe-vehicles, a smart device also collects and stores traffic data, road data and driver data in real time, however using a low-cost device to sense the environment around and inside the vehicle. In other words, we consider a smart device as a non-intrusive kind of sensor inside the vehicle and not embedded in it. Consider, for instance, smartphones, tablets or a hardware working as data sources inside a vehicle. In the following, we analyze proposals that rely on smart devices as Vehicular Sensor Data (VSD).

Aloul *et al.* [72] presented a smartphone app to detect and report car accidents automatically. They used accelerometer and GPS data to determine the severity of an accident and, if necessary, inform its location to the rescue personnel. Fox *et al.* [73] designed a crowdsourcing pothole detection scheme using real-world data collected from a smart device with sensors, such as GPS, vehicle speed, the three-axis acceleration and data from the mobility simulator CarSim [74]. They simulated an environment with 500 vehicles and were able to detect 99.6% of the potholes. In a real-world scenario, their approach could detect 88.9% of the potholes.

Goncalves *et al.* [75] designed a platform to acquire data about the traffic condition and to drive performance using a smartphone GPS. Han *et al.* [76] developed the SenSpeed, an accurate vehicle speed estimation system, to address an unavailable GPS signal or inaccurate data in urban environments. The authors relied on smartphone sensors, such as gyroscope and accelerometer to sense turns, stops and crossing irregular road surfaces. The results show that the real-time speed estimation error is 2.1 km/h, while the offline speed estimation error is 1.21 km/h, using the vehicle speed through the OBD as ground truth in their experiments. Ning *et al.* [77] conducted a study to detect traffic anomalies based on the analysis of trajectory data in Vehicular Social Networks (VSocN). Furthermore, they introduced a taxonomy for VSocN applications. The VSocN is an integration of social networks and the concept of the Internet of Vehicles (IoVs).

Chu *et al.* [78] designed a solution to distinguish driver and passengers based on accelerometer and gyroscope data of a smartphone. The Driver Detection System (DDS) focus on identifying micro-activities that can be discriminated using a popular and low-cost device. The results show an accuracy of up to 85% to determine who is the driver and the passenger. Aiming to identify the user's driving style, Vaiana *et al.* [79] used acceleration data (longitudinal and lateral) from a smartphone GPS. Kaplan *et al.* [80] reviewed and categorized techniques found in the literature for detecting driver drowsiness and distraction. They provided insights on techniques used for driver inattention monitoring and the recent solutions that use smart devices, such as smartphones and wearables.

Johnson and Trivedi [43] developed an inexpensive way to detect and recognize driving events and driving styles based on a smartphone. They created a MIROAD system that uses Dynamic Time Warping (DTW) and a smartphone equipped with a gyroscope, magnetometer, accelerometer, GPS and video recording capability to detect, recognize and record actions without external processing. The results proved that the MIROAD was able to recognize the U-turn 77% of the time. Similarly, however broader, Engelbrecht *et al.* [81] used accelerometer and gyroscope of a smartphone to recognize driving maneuvers. They validated the approach with an extra device equipped with a dedicated GPS and IMU. Hong *et al.* [50] created a model to identify an aggressive driving style. When using the smartphone and ECU data, they achieved an accuracy of 81%, while using only the smartphone the accuracy was of about 66.7%.

Fazeen *et al.* [82] also used smartphone sensors (three-axis accelerometer and GPS) to evaluate a vehicle's condition, such as gear shifts and road conditions (bumps, potholes, rough road, uneven road, and smooth road) and also various driver behavior. Paefgen *et al.* [44] conducted a study to evaluate driver behavior based on critical driving events and capture driver variability under real-world conditions. They compared the results of using only a smartphone and its inertial sensors to a commercial sensor unit [83] connected directly to the vehicle's OBD port. Castignani *et al.* [84] analyzed the capability of smartphone sensors to identify driving maneuvers and classify them as calm and aggressive. For such purpose, they developed the SenseFleet application. They used GPS and motion sensors from the smart device and also the weather and time of day to give them context information. They showed that SenseFleet can provide accurate detection of driving risks.

Yuan *et al.* [85] proposed the AC-Sense, an adaptive and comprehensive scheme for data acquisition in VSNs aiming to increase the quality of vehicular sensing. They used real taxi GPS trajectories and air quality data from Beijing. They combined these datasets to determine the capacity of taxis to sense the air quality. The results showed that the scheme can increase the sensing efficiency and maintain the data quality. Pan *et al.* [86] also used real taxi GPS trajectories to detect and describe traffic anomalies. Wang *et al.* [87] used vehicular trajectories with the location, heading and speed information to estimate the urban traffic congestion and detect anomalies on the road.

Bergasa *et al.* [88] developed a smartphone app to detect the safety level while driving. The app, DriveSafe, was developed for iPhone and aimed to detect inattentive driving behaviors, alerting the drivers about unsafe behaviors. To achieve that goal, the authors relied on computer vision and pattern recognition techniques and data from the rear camera of the smartphone, microphone, inertial sensors and GPS. They also presented a general architecture of DriveSafe and evaluated its performance in a testbed using data from 12 participants (9 males and 3 females). Each participant carried out two types of tests (aggressive and normal). The tests involved 20 minutes of trips during different days and times. DriveSafe was able to detect an inattentive driver behavior with an overall precision of about 92%. They also compared DriveSafe to the commercial AXA Drive app [89] and obtained better results.

Ma *et al.* [90] proposed the DrivingSense, which uses noise and other types of data provided by smartphone sensors to identify dangerous behaviors, such as speeding, irregular driving direction change and abnormal speed control. DrivingSense was able to detect events like driving direction changes and abnormal speed with a precision of 93.95% and 90.54%, respectively. Saiprasert *et al.* [91] also proposed algorithms to detect and classify driving events based on smartphone sensors, such as GPS and accelerometer.

Magaña and Muñoz-Organero [48] used location and road slope data obtained using a smart device to determine the risk of an accident based on the location of trajectories. Paefgen [67] focused on the automobile insurance market to introduce an adaptive insurance tariff known as PAYD. Bröring *et al.* [37] developed an app (EnviroCar) for Android smartphones to collect the location of vehicles and upload it to the Web.

Zhang *et al.* [46] developed a model to classify dangerous drivers using only smartphone sensors like accelerometer, gyroscope and GPS. The classification model obtained an accuracy of about 79.88% in a controlled environment and 80.00% in a naturalistic environment. Araújo *et al.* [65] developed an application to assess the driving behavior and reduce the fuel consumption. For that, besides in-vehicle sensors, they also used an accelerometer and GPS from a smartphone to acquire acceleration, altitude and location data.

Some studies [38], [49], [61] rely solely on the smartphone GPS to develop an app to help drivers improve their driving behavior. AbuAli [45] used GPS data to track the vehicle's location and store it on the Web. Rutty *et al.* [58], [60] also used a GPS provided by CarChip [59]. Riener and Reder [54] developed a social driving app aiming to improve the driving efficiency by providing recommendations about how to drive on a given track. Besides using in-vehicle data, they also relied on smartphone GPS and 3-axis acceleration data. Wang *et al.* [92] developed a system for visually analyzing urban traffic congestion. They used GPS trajectories and speed data from taxis in Beijing to design a model to extract and derive traffic jam information in a realistic road network. The process consists of an efficient data filtering step based on spatiotemporal aspects, size and network topology to create a graph structure and its visualizations.

B. Extra-Vehicular Sensor

The Extra-Vehicular Sensor (EVS) concepts of VDS corresponds to the subset of real and virtual sensors that seek to describe the driver's behavior and the environment around the vehicle by a variety of sources individually or fused. In that way, we categorize studies that use Questionnaire as Vehicular Sensor (QVS), Infrastructure as Vehicular Sensor (InfraVS) and Media as Vehicular Sensor (MVS), to provide data such as a descriptive driver's style, traffic behavior, weather conditions, and statistics related to drivers, gender, number of accidents, injuries, fatalities and others. In the following, we analyze each category and the related work.

1) Questionnaire as Vehicular Sensor: QVS can be considered the first way to sense the driver's perception of the road condition, accidents, distractions, behavior, expertise, feelings, gender and social aspects. Despite the high cost of applying a questionnaire, it gives a very detailed information about the context evaluated. There are studies that use a very known questionnaire of the Psychology to evaluate the aforementioned issues.

Driving involves a variety of skills including cognitive aspects such as attention and perception, but also emotional, motivational and social interaction. In that direction, the way in which a person performs this activity is described as the driving style. Moreover, it is well known that the driving style can lead to an inattentive and distracted direction, representing a significant issue to the road safety. There are different ways of understanding the driving style of a person or group. A wide solution adopted by psychologists is a questionnaire. There are diverse measurement instruments designed for this purpose, as the Driving Behavior Questionnaire [93], Driving Behavior Inventory [94], Driving Style Questionnaire [95] and Driving Expectancy Questionnaire [96].

Beanland *et al.* [97] conducted a study to identify driver distraction and inattention in serious crashes, based on the Australian National Crash In-depth Study (ANCIS). The participation in ANCIS was voluntarily and represents a person who was admitted to a hospital for getting involved in an accident. The authors indicated that the most severe injury accidents involve driver's inattention. Despite the variety of observed inattention and distraction events, most of them are possible to prevent. The development of interventions to the driving style depends on studies about the driving behavior and personality traits. In that direction, Poó and Ledesma [98] used the Zuckerman-Kuhlman Personality Questionnaire [99] to assess the relationships among driving styles and personality traits, and their variation by gender and age. As result, they showed a more comprehensive understanding of personality traits and driving style relationships. Hong *et al.* [50] obtained 81% accuracy in their method to determine the aggressiveness of driving style, using Manchester Driving Behavior Questionnaire (DBQ).

Van Huysduynen *et al.* [100] validated the different factors of Multidimensional Driving Style Inventory (MDSI) [101], aiming to know if the questionnaire can measure driving styles. Also, they grouped the factor analysis in angry driving, anxious driving, dissociative driving, distress-reduction driving and careful driving style. Sagberg *et al.* [102] conducted a vast literature review, aiming to understand the multidimensionality and complexity of driving styles. They found evidence that sociocultural factors, gender, age, driving experience, personality, cognitive style, group and organization values, and culture can determinate the driving style. The authors also observed the correlation between self-report instruments and observed behavior methods. Finally, but not limited to, they proposed a framework for predictions about how driving styles are established and modified, creating a base to test future empirical studies.

Truxillo *et al.* [103] developed a study to compare the effectiveness of the supervisor training and the use of eco-driving educational materials to reduce the fuel consumption. They collected data through a survey, containing the attitudes, knowledge and behavior of a driver before using eco-driving educational materials. After that, they disseminated the material to those participating organizations and the second and third surveys were sent after two and four months, respectively. As part of their results, they found that both groups increased the eco-driving behavior, suggesting that the support for efficient driving behavior can change the fuel consumption.

2) Infrastructure as Vehicular Sensor: The infrastructure can also tell about the vehicle's state, traffic condition, weather and driver's behavior. The essential difference compared to the IVS way is that the InfraVS also can provide information about the group and not only the vehicle individually. Although, this kind of vehicular sensor shows information at different granularity compared to the IVS. The infrastructure gives an external and global view of the environment, in this case, the transportation view. In the following, we describe approaches that use infrastructure data to develop or evaluate, somehow, the proposed applications.

Aoude *et al.* [104] developed algorithms for estimating the driver's behavior at road intersections. They used a set of devices that provide data for the further analyses, as GPS to record the current time of each vehicle, four radars which identified the vehicles, their speed, range and lateral position, four cameras, and a phase-sniffer to record the traffic light signal phase. The authors introduced two classes of algorithms that can classify drivers as compliant or violating. Finally, their approach was validated using naturalistic intersection data, collected through the U.S. Department of Transportation Cooperative Intersection Collision Avoidance System for Violations (CICAS-V).

Castignani *et al.* [84], in contrast to the current solutions, used contextual information, weather condition [105], in their application SenseFleet, aiming to better describe the driving behavior. Yuan *et al.* [85] used the air quality data in Beijing to create the AC-Sense, an adaptive and comprehensive scheme for data acquisition in VSNs. Wang *et al.* [87] proposed a traffic congestion detection based on GPS trajectories, social media and infrastructure data (e.g., weather), and showed that it could affect traffic conditions, leading to complementary traffic information.

Lu *et al.* [22] discussed the challenges and review the state-of-the-art about wireless solutions for vehicle communication among different entities, as vehicle-to-sensor, vehicle-to-vehicle, vehicle-to-Internet and vehicle-to-infrastructure. Using VISSIM traffic simulator [64], Zhao *et al.* [63] proposed and evaluated the Dynamic Traffic Signal Timing Optimization Strategy (DTSTOS), aiming to reduce the total fuel consumption and traffic delays in a road intersection, based on the vehicle speed, fuel consumption and traffic light timing control.

3) Media as Vehicular Sensor: Nowadays, with the growing and popularity of the Internet, the use of media to report the transportation conditions has increased. Thus, issues as incidents, traffic conditions, fatalities, road condition and events in a given location become the goal of different media platforms. We consider MVS as any kind of media (e.g., social media, blogs, news, map tools with transit insights, and government reports) that disseminate information to better contribute to transportation comprehension. The highlight is the social media data with the potential to be used as a real-time traffic data source. In the following, we describe approaches that use some sort of media data to develop or evaluate the proposed applications.

Pan *et al.* [86] proposed a method to detect and describe traffic anomalies based on GPS from vehicles' trajectories and social media data. The system provides real-time alerts when anomalies are detected, including the associated features and an event description based on social media. They used a GPS trajectory dataset of taxis to detect anomalies and the Twitter to provide details of these events. As result, the system detected 86.7% of the incidents reported to the transportation authority, whereas the baseline reported only 46.7%. Santos *et al.* [106] argued that LBSM feeds may offer a new layer to improve traffic and transit comprehension. They presented the Twitter MAPS (T-MAPS), a low-cost spatiotemporal model to improve the description of traffic conditions through tweets. The authors developed three route description services based on natural language analyses, aiming to enhance the route information. In the same direction Rettore *et al.* [107] developed the Twitter Incident (T-Incident), a low-cost learning-based road incident detection and enrichment approach built using heterogeneous data fusion. They designed a spatiotemporal grouping approach to fuse incident data, not-incident data, and LBSM data. Next, feeding a learning-based model to identify text patterns to detect the event types, achieving scores above 90%.

Gu *et al.* [108] explored the posts from the Twitter platform to extract traffic incident information, which is a low-cost solution compared to existing data sources. In that way, the authors developed a methodology to data acquisition, processing and filtering. They validated the Twitter-based incidents using data from RCRS (Road Condition Report System) incident, 911 Call For Service (CFS) incident, and HERE travel time (a part of the National Performance Management Research Data Set). That study pointed out the significance of traffic incident reported by Influential Users (IU) and individual users, frequency of reports on weekends and weekdays, and also during the day, and the volume of information from the center of a city and outside it. As conclusion, they demonstrated the potential of social media data to enrich the incident reporting sources.

In the same way, but using different social media as a data source, Septiana *et al.* [109] used text mining system about RSS feed Facebook E100 aiming to categorize road conditions into six types: floods, traffic jams, congested roads, road damage, accidents and landslides. They showed an accuracy of 92% in the road condition monitoring. Shekhar *et al.* [110] focused on the vehicular traffic monitoring using more than one social media, instead of traditional traffic sensors and satellite information which can be quite expensive. Using a Natural Language Processing (NLP) technique, they examined Twitter and Facebook posts to address traffic problems at a specific location and time interval. Besides, they looked for the causes of recurrent traffic congestion, and noticed that the obtained results were consistent when compared to the HERE Driver+, since more information was added to the context analysis.

Wang *et al.* [87] proposed a framework to integrate GPS trajectories data and social media data, aiming to compute urban traffic congestion more precisely. Using vehicular trajectories with location, heading and speed, social events from Twitter, road features, Point of Interest (POI), and weather information, they estimated the urban traffic congestion and also detected anomalies on the road. Sinha *et al.* [111] discussed the management of urban infrastructure based on

TABLE II
SUMMARIZING OF DATA SOURCE IN VEHICULAR DATA SPACE TAXONOMY

Papers	Vehicular Data Space: A Source Point of View					
	Intra-Vehicular Sensor			Extra-Vehicular Sensor		
	ECU	Probe-Vehicle	Smart Device	Infrastructure	Questionnaire	Media
[51], [52], [57], [56] [62]	✓					
[68], [69], [70]		✓				
[92], [82], [75], [81] [78], [79], [76], [88] [72], [73], [80], [90] [77], [91]			✓			
[39], [47], [66], [40] [41], [42]	✓	✓				
[43], [65], [44], [49] [67], [54], [37], [38] [58], [60], [45], [46] [48], [61]	✓		✓			
[104]				✓		
[98], [97], [100], [102] [103]					✓	
[50]	✓		✓		✓	
[84], [85]			✓	✓		
[22]	✓		✓	✓		
[63]	✓			✓		
[87]			✓	✓		✓
[107]				✓		✓
[108], [109], [110], [111] [112], [113], [106]						✓

insights from public data, which was used to categorize and visualize the urban public transportation issues. Their holistic framework considered the public transportation agency data, social media as Twitter and Facebook posts, and Web portals. Their goal was to help governments and common citizens to have a whole visualization and understanding of transportation in a city. Kurkcu *et al.* [112] proposed to fuse data from the Transportation Operations Coordinating Committee (TRANSCOM) and Twitter posts to allow real-time, inexpensive and geographical coverage. Using Twitter and Sina Weibo, Lau [113] presented an approach to extract and analyze traffic information to enhance ITSs.

C. Considerations

As previously mentioned, in this section, we discussed the studies considering the Vehicular Data Space. Table II summarizes recent proposals and their respective categories based on our taxonomy.

This area provides some initial and exciting results that can lead to new research challenges, when considering the data aspects and their applicability, as discussed in Section VI.

It is interesting to observe that there are studies in Vehicular Data Source (VDS) that considered a different number of data sources in their proposals. In particular, for one data source we have the following: Engine Control Unit; probe-vehicles; smart devices; infrastructure; questionnaire, and some sort of media. Considering the intersections of data sources, there are studies that used simultaneously two data sources: Engine Control Unit and probe-vehicles; Engine Control Unit and smart devices; Engine Control Unit and infrastructure; and smart devices and infrastructure. For three data sources, we have: Engine Control Unit, probe-vehicles and questionnaires; Engine Control Unit, smart devices and infrastructure; and smart devices, infrastructure and media.

Additionally, we can quantify the use of each data source in the studies above. Figure 7a shows the percentage of the

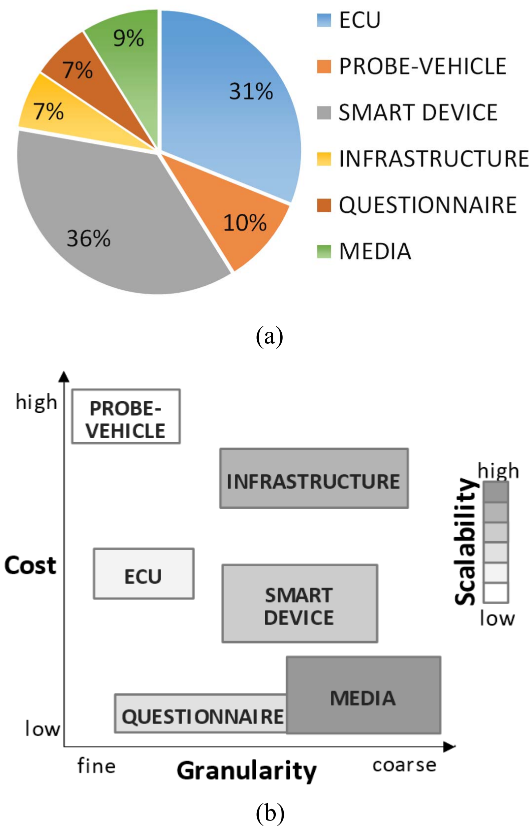


Fig. 7. (a) Most used data source in VDS. (b) An overview of data acquisition based on its granularity and financial costs.

use of each vehicular data source. Smart Device (typically smartphones) and ECU represent approximately two-thirds of all data sources employed in the development of applications and methods for ITSs. Smartphones are being designed with more and more sensors capable of sensing different physical variables, which explain their large use as a data source. An ECU also allows to sense the environment with high-quality sensors and assess the driver's behavior.

Next comes the Probe-Vehicle data source. In this case, only active research groups and companies use this data source due to its high cost to equip the vehicle and design solutions based on the embedded technologies.

The three least used data sources are Media, Questionnaire and Infrastructure. The use of media as a data source to the ITS has increased in the last years, and, probably, we can expect a stronger presence in the future. Media has the power to overcome the limitations of the data coverage provided by all other data sources mentioned in this study. Moreover, media can also offer the transportation view through the lens of users, companies and governments. Questionnaires report the behavior of a group and depend on the sample, and, thus, cannot be generalized. We noticed that the investigations about ITS do not use too much this data source such as media and its variations. Finally, but not less important, the infrastructure has taken its initial steps to be a data source to the VDS. The reasons are the low incentive, security and privacy issues to make the data available to the community.

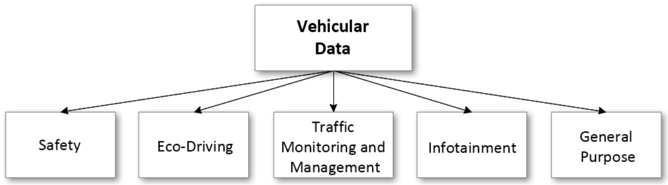


Fig. 8. Applications based on vehicular data.

While these issues keep untreated, we have to live with a short range of data, conducting studies only in large cities, which know the importance of having data available to investigate new applications and services to their citizens.

Figure 7b shows the relationship between *Costs* to develop and use of each VDSource and its respective *Granularity* and *Scalability*.[11] *Cost* represents the value to use a data source, *Granularity* how much descriptive the data source can be, and *Scalability* the capacity of acquiring large amounts of data from different agents.

The questionnaire is one of the cheapest ways to acquire vehicular data. However, their responses may not completely correlate with real-world events. On the other hand, the use of infrastructure as vehicular data is more scalable given its capacity to sense a variety of agents[12] in the transportation system. However, it typically involves high financial costs and a management solution for the transportation system. An example of a low-cost and scalable solution to acquire vehicular data is the use of social media as a vehicular data source. Its broad use allows a wide information dissemination about road conditions, accidents and other events.

TABLE III
CLASS OF DATA FROM VDS BASED ON A GIVEN APPLICATION GROUP

Application Group	Goals	Authors	Vehicular Data Space	
			Intra-Vehicle Data	Extra-Vehicle Data
Traffic Monitoring and Management	Event Detection (Incidents, Potholes, Traffic)	[68], [86], [63] [87]	✓	✓
		[92], [75], [76]	✓	
		[108], [109], [110] [112], [113], [111] [106], [107]		✓
Safety	Driver Style/ Behavior	[104], [50], [84]	✓	✓
		[114], [43], [44] [44], [82], [49] [39], [69], [81] [79], [78], [88] [70], [41], [52] [42], [46], [51] [90], [91], [56]	✓	
		[97], [98], [100] [102]		✓
	Event Detection (Incidents, Potholes, Traffic)	[72], [73], [40] [49], [54], [45] [77]	✓	
	Insurance, Fleet Monitoring, Aftermarket	[59], [83], [115]	✓	
Eco-Driving	Driver Style	[65], [116]	✓	✓
		[66], [62], [49] [58], [60], [61] [57]	✓	
		[103]		✓
	Event Detection (Incidents, Potholes, Traffic)	[48], [63]	✓	✓
		[54]	✓	
	Data Acquisition	[37]	✓	
General Purpose	Data Acquisition, Data Available, Developers	[38], [85]	✓	✓
		[114], [88], [37] [55], [117], [118] [119]	✓	

Intra-Vehicle Data = Location, Speed, RPM, Acceleration, Brake Pedal, Engine Load, Throttle Position, Gear, Fuel, Emissions, Engine Temp, Turning, Radar, Video/Audio,Light; Extra-Vehicle Data = Altitude/Atmospheric Pressure, wind speed/humidity/temperature, and traffic light/inductive loop; sociocultural factors, gender/age,driving experience, personality, and cognitive style; Social Media, News, and Government data;

V. POTENTIAL APPLICATIONS

Many are the applications designed for the vehicular environment, with different functions and goals. In this section, we categorize these applications based on the taxonomy described in Section IV. Figure 8 depicts the main applications based on vehicular data related to safety, eco-driving, traffic monitoring and management, infotainment, and also general purpose.

To present an overview of the applications, we summarized them in data classes using the VDS. We grouped the investigations into two data categories: Intra-Vehicular Data (IVD) and Extra-Vehicular Data (EVD). Table III describes the groups of applications mentioned before. We noticed that 64% and 16% of them only used Intra-Vehicular Data (IVD) and Extra-Vehicular Data (EVD) to develop their applications, respectively, whereas 20% dealt with both groups. This clearly shows some interesting opportunities to explore the EVD and the fusion between IVD and EVD. For the rest of this section, we discussed the data of each category used by a given investigation. Furthermore, we highlighted the data availability which most of those group of applications utilized, and overview the whole section at the end.

[11]It means the capacity to provide amounts of data from a variety of agents.
[12]For instance, people, vehicles and companies.

A. Safety

There are many ways to increase the safety on the roads. The advance of technology has allowed investments on vehicles and roads to achieve this goal. Some studies support the necessity of improvements to decrease the number of road accidents. Most accidents could be avoided if the driver received a warning half a second before the moment of collision. In that way, studies to improve the recognition of driver's style have emerged, aiming to better understand the driver's behavior. In the safety category, we considered applications that propose to identify driver's patterns (e.g., style, behavior), offer customized insurance services, and improve the car security.

Driving analysis is a topic of interest due to the increase of the safety issue in vehicles. In 2015, the U.S. Department of Transportation showed the number of deaths in motor vehicle crashes, which is above 35 thousand people [120]. They also argued that alcohol, speeding, lack of safety belt use and other problematic driver's behaviors contribute to the death in vehicle crashes. The driver's behaviors vary considerably depending on age, gender, drugs consumption, types of used roads, distracted driving attitudes [121], and other factors. For these reasons, the study of driver's style has emerged, aiming to increase driving safety and, as consequently, reduce deaths in traffic. Engelbrecht *et al.* [25] analyzed the use of smartphones to support a variety of ITS applications in a safety field as the driver's behavior, and road condition monitoring. Kaplan *et al.* [80] also conducted a review to detect driver's drowsiness and distraction.

Considering as input data acceleration, braking and turning collected from the accelerometer sensor of a smartphone, once inside the vehicle, it is possible to sense the vehicle longitudinal and lateral acceleration. Then, thresholds on these measurements can detect different maneuvers. In that way, if we apply thresholds on the z-axis (representing acceleration and brakes), we can obtain rules to define the driver's style, aiming to identify sharp peaks that indicate aggressive increases of speed or hard braking. Additionally, analyzing thresholds on the x-axis acceleration, it is possible to detect excessive speed in left or right turns.

Several studies have focused on driving style and driving maneuvers recognition [39], [41], [42], [43], [46], [49], [51], [52], [54], [56], [79], [81], [82], [84], [88], [91]. Some of these studies identify who the driver is whereas others classify the driver's behavior as aggressive or normal, and driving maneuvers. Ma *et al.* [90] discussed the influence of noise provided by smartphone sensors, to identify dangerous behaviors. Satzoda and Trivedi [47] extracted semantic information from raw data provided by the vehicle. D'Agostino *et al.* and AbuAli [40], [45] proposed a classification method for driving events recognition, using short-scale driving patterns. Fox *et al.* [73] designed a pothole detection scheme using a real-world data and simulator.

In the same way, Aoude *et al.* [104] developed algorithms for estimating the driver's behavior at road intersections. Wang *et al.* [122] presented a survey of a wide range of mathematical identification and modeling methods of driver's behavior. Guo and Fang [69] conducted a study aiming to identify factors associated with individual driver's risk and also predict the high-risk drivers, based on demographic data, driver's personality, and driving characteristics. Elhenawy *et al.* [70] presented a model that can be integrated with in-vehicle safety systems to predict driver's stop/run behavior and then taking actions to avoid collisions. Chu *et al.* [78] developed a smartphone app that focuses on determining if its user is a passenger or a driver. Using different approach, [97], [98], [100], [102] used questionnaires from the literature to understand the multidimensionality and complexity of the driving styles concept. Hong *et al.* [50] developed a platform, aiming to model the aggressiveness of the driving style, based on different data sources as smart devices, ECU and questionnaire.

Another agent interested in issues related to the vehicle safety is the manufacturers. They pay attention to their vehicles' behavior to foresee problems, allowing them to offer their services in advance. Thus, in that class of application, the manufacturers use the vehicular sensor data to improve their technology to make their automobiles safety and comfortable. As safety applications, we have other two classes as prevention and correction. A diagnostics application is included in the prevention class and provides information about the components malfunction, aiming to avoid further breakdowns or damages. The applications in the correction class is designed to protect the vehicle and its passengers. The airbag application is activated based on a sudden stop (in most cases), the wheel speed can be changed depending on the lack of traction, for instance.

Many approaches considered the high costs involved in evaluating and improving vehicular safety solutions. They allowed a low-cost way for companies and researchers to develop and test their solutions. As an example, CarSim [74] or generally VehicleSim (VS) is a product conceived to provide a realistic view of the vehicle components (e.g., tires, suspension, and steering) in different environments. Many companies and researchers use it as a tool for kinematic and control simulation testing to improve their development process.

Other market solutions focused on fleet companies. For instance, the CarChip Connect [59] is an easy-to-use fleet monitoring tool. CarChip is a small telematics device with GPS and accelerometer, which connects to the vehicle by the OBD-II port. This tool provides the vehicle location and real-time alerts to improve the safety and the productivity. This tool tracks and sends reports data to the cloud, allowing clients to manage their fleets. In the same way, Scope Technology [83] aims to provide end-to-end telematics products and services. Their solutions empower insurance providers, fleet operators and aftermarket service providers to implement their personalized services.

The possibility to sense the vehicle and detect the driver's behavior opened the opportunity to customize applications and services developed according to the client's needs. As an example of these approaches, there are applications for insurance companies aiming to offer personalized services to their customers. The concepts of PAYD or Pay-How-You-Drive (PHYD) promote a new vision of how to charge rates, based not on the range of risk as age, address and gender, but also considering the driver's behavior, i.e., aggressive or standard. The aim of these applications is to classify the drivers' behavior to describe a distinguished attitude and its respective degree of safety for themselves and all around them. Besides that, the ability to offer flexible insurance services promises a significant improvement in traffic safety, taking into account the incentive to customers to drive safely.

Paefgen [67] focused on evaluating an accident risk based on continuous measurement of vehicular sensor data in the context of adaptive insurance tariffs. That work of Telematics strategy for automobile insurers also pointed out the business implications of risk-adaptive insurance taxes. Showing the less applicability to the current market, but a promising perspective on the new market entrants. As an example of a market, AXA is an insurance company that focuses on protecting personal property (e.g., cars, homes) and liability (personal or professional). AXA Drive [89] gives the driver real insights and personalized tips to help them to improve their driving behavior. State Farm insurance company developed a smartphone app, Drive Safe & Save [123], aiming to offer to their clients the reduction of auto insurance based on safer driving. Besides the car insurance, another promising field is related to the Health insurance. It aims to provide fast medical assistance based on a smart device application that automatically detects serious vehicle crashes, also known as Real-Time Medical Response [124].

TABLE IV
VEHICULAR DATA SPACE FOCUS ON SAFETY APPLICATIONS

Application Group	Goals	Authors	Vehicular Data Space																			
			Location	Speed	RPM	Acceleration*	Brake Pedal	Engine Load	Throttle Position	Gear	Fuel	Emissions	Engine Temp*	Turning*	ATM*	Radar	Video/Audio	Light	Infrastructure	Questionnaire*	Media*	Car Features
Safety	Driver Style/Behavior	[39]	✓	✓		✓	✓							✓			✓					
		[41]	✓	✓	✓	✓	✓		✓					✓								
		[43], [88], [90]	✓			✓								✓			✓					
		[44]	✓	✓		✓	✓							✓								
		[46]	✓		✓	✓		✓	✓					✓								
		[51]		✓	✓		✓		✓					✓								
		[69]	✓	✓		✓										✓	✓					
		[70]	✓			✓	✓							✓			✓					
		[78], [81]				✓								✓								
		[79]				✓																
		[82], [91]	✓			✓																
		[44]	✓			✓								✓								
		[97], [98], [100], [102]																		✓		
		[104]	✓													✓	✓		✓			
		[114]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓					
		[52]			✓		✓		✓													
		[42]		✓	✓												✓					
		[49]	✓	✓	✓	✓																
		[50]	✓	✓	✓	✓			✓					✓						✓		
		[84]	✓	✓		✓								✓					✓			
		[56]		✓	✓			✓	✓	✓	✓	✓										
	Count Result	**28**	**17**	**11**	**9**	**19**	**7**	**3**	**7**	**2**	**2**	**2**	**1**	**15**	**0**	**3**	**9**	**0**	**2**	**5**	**0**	**0**
	Event Detection (Incidents, Potholes, Traffic)	[72]	✓			✓																
		[73]	✓	✓		✓																
		[40]	✓	✓		✓	✓	✓									✓					
		[49]	✓	✓	✓	✓																
		[54]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓					
		[45]	✓	✓	✓				✓													
		[77]	✓																			
	Count Result	**7**	**7**	**5**	**3**	**5**	**2**	**2**	**2**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**2**	**0**	**0**	**0**	**0**	**0**
	Insurance,Fleet Monitoring, Aftermarket	[59], [83], [115]	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓							
	Count Result	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**3**	**0**	**3**	**3**	**3**	**3**	**3**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Acceleration = longitudinal/3-axis, Engine Temp = Engine Coolant Temp, Turning = Rotation Angle;
ATM = Altitude/Atmospheric Pressure, Infrastructure = wind speed/humidity/temperature, and traffic light/inductive loop;
Questionnaire = sociocultural factors, gender/age,driving experience, personality, and cognitive style;
Media = Social Media, News, and Government;

Aloul *et al.* [72] also conducted a study in that way, with the development of a smartphone app to detect and report car accidents.

Section IV reviewed the literature through the lens of VDS and its data sources. However, we can have new insights when we look at the data used to achieve specific goals. Thus, Table IV classifies applications into three groups: (i) safety; (ii) application goals as driver style/behavior, event detection, and insurance, fleet monitoring, and aftermarket; and (iii) data used for these applications.

That table categorizes 38 applications as safety, of which 28 focused on the driver's style/behavior, 7 on event detection, and 7 on insurance, fleet monitoring and aftermarket. Section VI discusses the challenges and open issues associated to them.

B. Eco-Driving

Fuel consumption is a factor that varies according to the drivers' habits. Two different vehicles are expected to consume more or less fuel according to their engines' size. However, the same vehicle may behave differently depending on the person who is driving it. As an example, someone who drives a car aggressively and accelerates it more than another person who uses it more consciously is expected to consume more fuel. From both environmental and economic points of view, it is desirable that drivers interact with their vehicles in a way that is as fuel efficient as possible, which reduces costs with refueling and greenhouse gases emissions. Collecting vehicular fuel consumption and emission data can lead to applications that help drivers to optimize these aspects in their driving styles.

TABLE V
VEHICULAR DATA SPACE FOCUS ON ECO-DRIVING APPLICATIONS

Application Group	Goals	Authors	Location	Speed	RPM	Acceleration*	Brake Pedal	Engine Load	Throttle Position	Gear	Fuel	Emissions	Engine Temp*	Turning*	ATM*	Radar	Video/Audio	Light	Infrastructure	Questionnaire*	Media*	Car Features
Eco-Driving	Driver Style	[62]		✓	✓	✓		✓	✓	✓	✓	✓	✓									
		[57]		✓	✓			✓		✓	✓	✓										
		[103]																		✓		
		[116]			✓	✓					✓								✓			✓
		[65]	✓	✓	✓	✓			✓		✓				✓							
		[49]	✓	✓	✓	✓																
		[66]	✓	✓	✓		✓		✓		✓											
		[58], [60], [61]	✓	✓							✓	✓										✓
	Count Result	**10**	**6**	**8**	**6**	**4**	**1**	**2**	**3**	**2**	**8**	**5**	**1**	**0**	**1**	**0**	**0**	**0**	**1**	**1**	**0**	**4**
	Event Detection (Incidents, Potholes, Traffic)	[48]	✓	✓	✓	✓					✓				✓			✓				
		[63]		✓							✓								✓			✓
		[54]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓					
	Count Result	**3**	**2**	**3**	**2**	**2**	**1**	**1**	**1**	**1**	**3**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**0**	**0**	**1**
	Data Acquisition	[37]	✓	✓	✓			✓	✓		✓	✓	✓									✓
	Count Result	**1**	**1**	**1**	**1**	**0**	**0**	**1**	**1**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**

Acceleration = longitudinal/3-axis, Engine Temp = Engine Coolant Temp, Turning = Rotation Angle;
ATM = Altitude/Atmospheric Pressure, Infrastructure = wind speed/humidity/temperature, and traffic light/inductive loop;
Questionnaire = sociocultural factors, gender/age,driving experience, personality, and cognitive style;
Media = Social Media, News, and Government;

Different initiatives and studies [48], [49], [54], [57], [58], [61], [65], [66], [103] are investing specialized services for Eco-driving to encourage driving style improvements, in order to reduce fuel consumption. Eco-driving refers to behavior and techniques designed to reduce fuel consumption, which includes recommendations for a person's driving style, the way, and frequency they use a vehicle, its configuration, accessories and maintenance. Eco-driving is part of a comprehensive approach to reduce the transport sector's contribution to the greenhouse effect. Broring *et al.* [37] developed a solution to acquire vehicular data and made it available to the community.

Brace *et al.* [62] proposed a DAS to reduce fuel consumption decreasing the rates of acceleration, and the early gear changes, demonstrating a fuel savings of up to 12%, and average fuel savings of 7.6%. The CGI Group Inc [116] conducted a study based on more than 3 million Scania Truck trips, across seven European countries. They compared the impact of eco-driving coaching for different fleets and countries. Moreover, they proposed an estimated effect of coaching (EEOC), which provides a realistic estimate of the fuel savings gained from eco-driving coaching. Zhao *et al.* [63] proposed the dynamic traffic signal timing optimization strategy (DTSTOS), also aiming to reduce the vehicle fuel consumption in a road intersection.

Table V summarizes all applications reviewed in this section, grouping them in the following groups: (i) eco-driving application; (ii) application goals as driver style/behavior, event detection, and data acquisition; and (iii) data used for these applications. Thus, we categorized 14 applications as eco-driving, of which 10 focused on the driver's style/behavior, 3 on event detection, and 1 on data acquisition.

C. Traffic Monitoring and Management

It is well known the issues related to transportation and traffic in large cities, such as time spent on traffic jams, and number of fatalities and injuries on the roads, which achieved an alarming scenario. These numbers prompted new initiatives from governments and private sectors to improve the road traffic efficiency and safety. Thus, an ITS becomes a way to find smart and low-cost solutions to improve decision-making and obtain rich traffic information. In this field, to acquire rich information about the traffic, we need to comprehend the environment such as weather condition, vehicle characteristics and the road condition as influencers to the driving style. Thus, we show some applications that are interested in the characterization of traffic and road conditions.

Goncalves *et al.* [75] used a smartphone GPS to study and characterize traffic and road conditions. They built the Iris Geographic Information System (GIS)-based platform using the smartphone Android on a client side and a server side for collect data by store, pre/post processing, analyze and

TABLE VI
VEHICULAR DATA SPACE FOCUS ON TRAFFIC MONITORING AND MANAGEMENT APPLICATIONS

Application Group	Goals	Authors	Vehicular Data Space: A Data Point of View																			
			Location	Speed	RPM	Acceleration*	Brake Pedal	Engine Load	Throttle Position	Gear	Fuel	Emissions	Engine Temp*	Turning*	ATM*	Radar	Video/Audio	Light	Infrastructure	Questionnaire*	Media*	Car Features
Traffic Monitoring and Management	Event Detection (Incidents, Potholes, Traffic Condition)	[68]				✓											✓		✓			
		[92], [75]	✓	✓																		
		[76]	✓			✓								✓								
		[86]	✓																		✓	
		[87]	✓																	✓	✓	
		[108], [109], [110], [111] [112], [113]																			✓	
		[63]		✓							✓								✓			✓
		[106]																			✓	
		[107]																	✓		✓	
	Count Result	**15**	**5**	**3**	**0**	**2**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**3**	**1**	**10**	**1**

Acceleration = longitudinal/3-axis, Engine Temp = Engine Coolant Temp, Turning = Rotation Angle;
ATM = Altitude/Atmospheric Pressure, Infrastructure = wind speed/humidity/temperature, and traffic light/inductive loop;
Questionnaire = sociocultural factors, gender/age,driving experience, personality, and cognitive style;
Media = Social Media, News, and Government;

manage the traffic condition. Wang *et al.* [92] developed a system for visual analysis of urban traffic congestion, using only GPS trajectories. Han *et al.* [76] developed the SenSpeed an accurate vehicle speed estimation system to urban environments. Ning *et al.* [77] studied the traffic anomaly detection based on trajectory data analysis in VSocN. Using a public data, Gu *et al.* [108] explored the Twitter platform, aiming to extract traffic incident through users posts, providing a low-cost solution to increase the road information.

Santos *et al.* [106] and Rettore *et al.* [107] argued that LBSM feeds may offer a new layer to improve traffic and transit comprehension. They presented the Twitter MAPS (T-MAPS) a low-cost spatiotemporal model to improve the description of traffic conditions through tweets, and the Twitter Incident (T-Incident), a low-cost learning-based road incident detection and enrichment approach built using heterogeneous data fusion.

Septiana *et al.* [109] proposed the categorization of the road conditions, based on text mining of Facebook feeds. In the same way, Shekhar *et al.* [110] focused on the vehicular traffic monitoring using Facebook and Twitter posts. Pan *et al.* [86] also used social media data to enrich the anomalies detection based on GPS from vehicles trajectories. Sinha *et al.* [111] and Lau [113] presented some insights based on public data to enrich urban public transportation and the ITS. Kurkcu *et al.* [112] provided detailed information about incidents, based on agencies and social media data.

On the other hand, Mednis *et al.* [68] proposed the CarMote, a dedicated hardware designed to monitor and create a detailed road map of the quality of the surface and weather. Zhao *et al.* [63] proposed the DTSTOS, also aiming to reduce the traffic delays in a road intersection. Aquino *et al.* [125], [126] propose a characterization of vehicles velocities to identify traffic behaviors using information theory.

Table VI summarizes these initiatives and studies into three groups: (i) traffic monitoring and management application; (ii) event detection as application goals; and (iii) data used for these applications. We categorized 15 applications focused on event detection (e.g., incidents, potholes and traffic condition).

D. General Purpose

The general purpose category shows studies to develop solutions to data acquisition and its availability to the community. Table VII summarizes the proposals in this category. For instance, Bröring *et al.* [37] proposed a solution to acquire vehicular data and made it available to the community, showing applications to fuel consumption and emissions. However, with these data in a large covered area, the possibilities exceed that initial purpose. An adaptive and comprehensive scheme for data acquisition in VSNs was proposed by Yuan *et al.* [85], opening a variety of applications based on these data.

A smartphone app DriveSafe is available on the Internet [88] to detect the level of safety while driving. Furthermore, these data can be used to understand the safety of the driver and the safety of the road or area as well. There are initiatives [55], [117], [118], [119] that made available vehicular sensor data, which allows the industry and research groups to develop their solutions. A prototype to provide vehicular data access through a website was developed by Reininger *et al.* [38], which allows access to the vehicle speed, RPM, fuel consumption, GPS and altitude, making possible to design a variety of applications based on these data. Another data source that can be used as a general purpose is an international collaboration between Japan, Italy, Singapore, Turkey, and the USA, UTDrive [114]. The aim was to develop a framework for building models of driver safety behavior. Moreover, they made the data collected available to the community, allowing the wide developing of applications.

E. Infotainment

Infotainment is a term used in the vehicular context to provide services to the driver and passengers, based on

TABLE VII
VEHICULAR DATA SPACE FOCUS ON GENERAL PURPOSE APPLICATIONS

Application Group	Goals	Authors	Vehicular Data Space																			
			Location	Speed	RPM	Acceleration*	Brake Pedal	Engine Load	Throttle Position	Gear	Fuel	Emissions	Engine Temp*	Turning*	ATM*	Radar	Video/Audio	Light	Infrastructure	Questionnaire*	Media*	Car Features
General Purpose	Data Acquisition Data Available Developers	[37]	✓	✓	✓			✓	✓		✓	✓	✓									✓
		[88]	✓			✓								✓			✓					
		[55], [117], [118], [119]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓					
		[114]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓					
		[38]	✓	✓	✓						✓				✓							
		[85]	✓																✓			
	Count Result	**9**	**9**	**7**	**7**	**6**	**5**	**6**	**6**	**5**	**7**	**6**	**6**	**6**	**1**	**1**	**6**	**0**	**1**	**0**	**0**	**1**

Acceleration = longitudinal/3-axis, Engine Temp = Engine Coolant Temp, Turning = Rotation Angle;
ATM = Altitude/Atmospheric Pressure, Infrastructure = wind speed/humidity/temperature, and traffic light/inductive loop;
Questionnaire = sociocultural factors, gender/age,driving experience, personality, and cognitive style;
Media = Social Media, News, and Government;

a combination of information and entertainment. A variety of applications can be developed to achieve this goal. For instance, it is common that drivers bringing their data in smartphones through apps, either local or on the cloud. However, when they are driving, the use of smart devices become a risk to themselves and other drivers. Furthermore, a traditional hands-free approach has limitations in several applications. In this way, it is convenient to think that the apps in a driver's smartphones can become useful through the dashboard display and multimedia kit inside the cars. Many companies and research groups are investing in solutions to better involve drivers and the environment around them. In the following, we describe some initiatives and studies in that way.

GM developed OnSart [127] a solution to maintain its costumers connected with their own cars. OnStar uses an integrated cellular service to connect the car to the Internet, allowing drivers and passengers to use the car audio interface to contact OnStar representatives for emergency services, vehicle diagnostics, and directions or personalized trip information. Moreover, GM costumers can use a smartphone app to take control of their vehicles, for instance, lock doors, send an alarm to locate it, find it on a map, send a trip to navigate through the GPS embedded in a car, and also monitor it along the time. Similarly, Audi offers the Audi Connect [128] to give drivers more control over their vehicles, maintaining them connected all the time to the Internet through the 4G-lite cellular network.

Some automakers have invested to provide customers with highly integrated connected experiences through connected in-vehicle infotainment systems to smartphone applications. To achieve this goal, automakers in partnership with other companies like Apple, Google, Pioneer, and Sony, for instance, have developed a way to create that connectivity environment. A recent initiative created by Ford, named Smart Device Link (SDL) [129], aims to enable existing smartphone applications to interface with vehicles. Through an open source community, using a standard set of protocols and messages that connect applications of a smartphone to a vehicle head unit. There are initiatives [55], [117], [118], [119] that allow industries and research groups to develop their solutions using an in-vehicle data and connectivity. Cheng *et al.* [130] analyzed communication protocols and their suitability for infotainment and safety services in VANETs.

Generally, these approaches aim to safely permit the user to interact with apps installed in their smartphones while driving, exhibit the results on the dashboard display and hear the audio via the car's speakers. Another important issue is related to the variety of car models, not being restricted to one brand or model. The applicability can be diverse, for instance, get directions, make calls, send and receive messages, navigate on the Internet using voice recognition, and listen to music. In that direction, Apple developed the CarPlay [131] solution for their customers. The Car Connectivity Consortium (CCC) developed the MirrorLink [117], which enables to establish a connection with a list of compatible cars, smartphones and apps. Toyota and BMW have also an infrastructure for the users of Toyota Touch 2 [132] and BMW ConnectedDrive [133], respectively.

F. Data Availability

An important issue in the initiatives and studies discussed above is the data availability. This can allow new investigations based on to use of such data. Table VIII summarizes the availability of a given data as follows.

(i) *Partially Public:* not all data is available to the general public. It can be delivered with a reduced sampling rate or a low-frequency rate, with specific features blocked, and also with some sort of noise; (ii) *Public:* data is available to the general public, with no restrictions; (iii) *Private:* data is only available for closed groups or people ready to pay to have full access. Most available VDS data are free for the public or partially accessible by them. On the other hand, there are datasets provided by private companies, governments or even research groups with restrict access to the general public.

It is possible to see the partial availability of *fuel* and *emissions* data due to restrictions of vehicle sensors' data access applied by some automakers. The access to the *infrastructure* data is also restricted to a set of sensors such as camera and road speed of reduced areas. The availability of *Social Media*

TABLE VIII
AVAILABILITY OF VEHICULAR DATA SPACE

Availability	Vehicular Data Space																			
	Location	Speed	RPM	Acceleration*	Brake Pedal	Engine Load	Throttle Position	Gear	Fuel	Emissions	Engine Temp*	Turning*	ATM*	Radar	Video/Audio	Light	Infrastructure	Questionnaire*	Media*	Car Features
Partially Public									✓	✓							✓		✓	✓
Public	✓	✓	✓	✓		✓	✓				✓		✓						✓	
Private					✓			✓				✓		✓	✓	✓	✓	✓	✓	

Acceleration = longitudinal/3-axis, Engine Temp = Engine Coolant Temp, Turning = Rotation Angle;
ATM = Altitude/Atmospheric Pressure, Infrastructure = wind speed/humidity/temperature, and traffic light/inductive loop;
Questionnaire = sociocultural factors, gender/age,driving experience, personality, and cognitive style;
Media = Social Media, News, and Government;

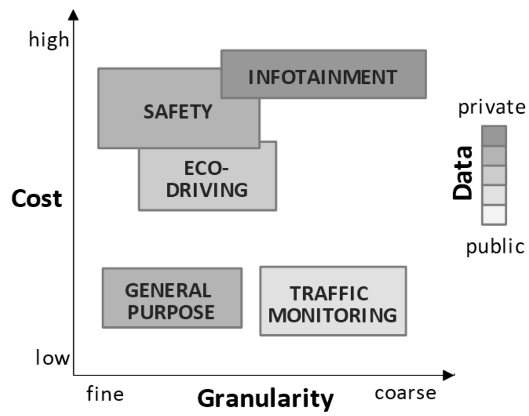


Fig. 9. Overview of application groups based on their granularity, financial costs and data availability.

data can be classified into three groups: full access; short sample of the dataset; and only paid access. Thus, initiatives and research groups that plan to use social media should be aware of these possibilities.

Based on the relationship between *Cost* and *Granularity* depicted in Figure 7b, and the *Data Availability* analysis, we evaluated each application group in terms of these three metrics. Figure 9 presents the cost and granularity, considering the data sources of a given data used by an application of VDS (IVD and EVD). Moreover, the evaluation of the data availability used by applications provides an access scale between *Public* and *Private* for a given application group.

We noticed that safety applications used fine-grained data to obtain high-quality results, but introduced a high cost due to the quality of the used sensors and the fact that datasets are non-public. Traffic monitoring applications typically have a reduced cost, given the use of low-cost sources and public data. However, these applications have to deal with coarse-grained data which may reduce their accuracy. Another important group of applications is the infotainment. The data availability related to that class becomes essential to provide personalized infotainment solutions to drivers and passengers. This will probably demand a thorough study to understand the drivers' behavior, traffic, consumption trends of information and products, among other issues. The associated costs will depend on the data granularity, quality and availability.

G. Overview

The safety application group described in Table IV reports 38 studies that list the most used data to detect the driver's behavior and events on the roads, but disregard the location and acceleration (longitudinal/3-axis). Driver's behavior applications also use the turning angle, which differs from the 3-axis acceleration due to its reduced noise. However this type of data comes from the ECU, and its access is not promptly available. On the other hand, IMU devices or smartphones can provide 3-axis acceleration, which provides a low-cost solution to detect the driver's behavior.

In event detection applications, locations play an essential role to identify the event on the map, and the speed and acceleration (longitudinal/3-axis) can provide semantic data from those locations. In insurance and fleet monitoring applications, there is a need for different sensor data, and possibly from smart devices as well, which will somehow identify the kind of behavior or status expected for the respective application.

It is important to notice that different data sources will have different roles in these applications (event detection, and insurance and fleet monitoring), and others as well. Sensor data such as fuel, emissions and light will possibly have no or little contribution to these previous applications. Social media data might be used to help identify the user's behavior and feelings, and, thus has the potential to be very useful in this case.

The eco-driving application group described in Table V reports 14 studies that list the most used data to detect the driver's style, events on the roads and evaluate an efficient fuel use, but disregard the location, speed and fuel consumption. This fact shows the intuitive relationship between vehicle speed and fuel consumption. Besides these data, RPM also contributes to these applications. The combined use of fuel consumption and brake pedal can offer a different solution for eco-driving applications. Social media and infrastructure can provide support for applications such as the shortest route, and

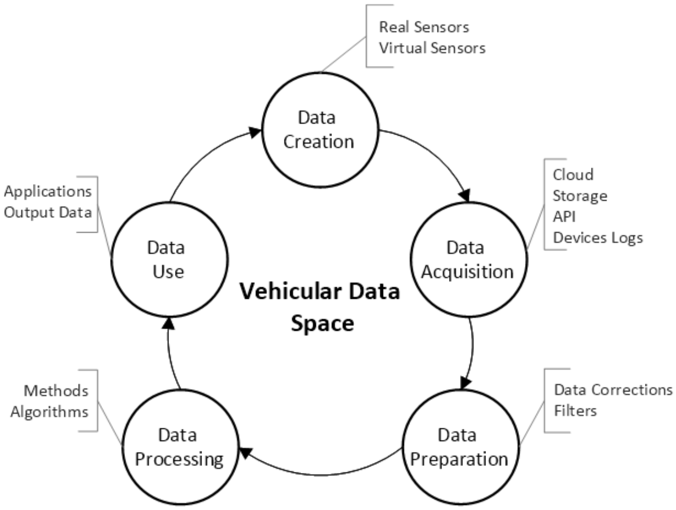


Fig. 10. The data cycle on the Vehicular Data Space (VDS).

near and cheapest gas station, which reduce emissions and fuel consumption.

We also observed that media data becomes an important data source in the traffic monitoring application group, where 10 of 15 studies use it to achieve their goals (see Table VI). This fact shows its capacity to describe events on the road from a user's perspective, which was not possible before. This is an opportunity to better manage the whole traffic and people's mobility.

Some studies showed the capacity of smartphones to measure movements and detect the driver's behavior. The comparison with the vehicle sensors from ECU is natural, making these smart devices an inexpensive way of instrumenting a vehicle. Moreover, smartphones have advanced sensors, allowing them to recognize the driving style, road and traffic conditions, and vehicle condition. On the other hand, there are substantial challenges involved in detecting movements using smartphones. The first one is the noise that comes from the vehicle movement and the uneven road. Besides, the position of the device can affect the results. Failures can occur considering that these devices are for general purpose. For instance, notifications of some applications can have a higher priority to the operating systems, and, then, the real-time measurement can be interrupted. Last but not least, real-time data is an essential feature for a driving analysis. However, continuous sensing and processing can drain the battery, making it impracticable for the users.

VI. Challenges and Open Issues

This section discusses the characteristics of VDS, which represent challenges and opportunities for future work according to the data cycle of the VDS depicted in Figure 10. The main issues can be classified into five stages as the process of *Data Creation*. The ways of *Data Acquisition*, its availability to the community, and its spatiotemporal coverage which constitutes limitations to developing general and broad solutions. Also, the issues related to the data storage and the data structure, that becomes relevant in the ITSs, due to needing of the big data analysis. The *Data Preparation* step, in general, represents the most critical stage of each work, where fit the data depict a crucial tread to the solution proposed. We also highlight the *Data Processing* to transform the treated data into valuable or more informative data absorbed by the next step, *Data Use*. This last step provides the application to the users, or also the output data to a new data cycle. Based on that, we deeply analyzed 53 of 62 papers (considering some limitations to identify each step) looking for this data cycle, aiming to show possibles challenges and open issues of each data stage explored in the literature.

A. Data Creation

In this stage, we highlight three relevant issues like Virtual Sensor, the Grow Up of Data Creation, and Electric Vehicle (EV), which deserve a more comprehensive discussion. The data cycle of the VDS begins with data creation. Data can come from real sensors responsible for measuring the environment or virtual sensors. In this stage, a problem that arises when using real sensor data to monitor and control entities, especially vehicles, is the data reliability, which includes availability and data quality. A sensor must generate correct readings continually, and control systems depend on these characteristics to operate properly. However, every sensor has an inherent probability of presenting a malfunction, which might affect these issues.

A solution to monitor and improve physical sensors, or temporarily replace them, is to use a virtual sensor. This type of sensor may combine data from other sensors, correct or filter failures, apply adequate methods and algorithms of a given problem domain, and take its data to an applications or input it to a new cycle. Virtual sensors are useful alternatives to monitor aspects, variables and phenomena for which there are no physical sensors. There are cases where physical sensors are unavailable, and a virtual sensor can replace them, given that the variable they monitor is correctly described or highly correlated to other monitored variable. In fact, a virtual sensor may substitute several physical sensors by combining their data using models and outputting the desired information. Additionally, a virtual sensor can produce new and higher level sensor information.

A relevant issue in VDS is the financial investment to create/obtain new data. In practice, the cost to embed sensors or smart devices in a vehicle and evaluate the experiments reduce the capability of initiatives to investigate a new and unexplored field. Besides, projects that developed their own data acquisition infrastructure [41], [46], [48], [60], [61], [63], [90], [91], for instance, do not make all data available to the community for some reasons: (i) vehicular data has commercial value and may not be available at all; (ii) those initiatives keep their propriety to advance their investigations and innovations; and (iii) data privacy may become a problem in case of releasing the datasets. These are facts that help us to understand the small availability of vehicular data, which reflects in just a few investigations [51], [52], [85], [87], [108], [110], [112], [113] that use data provided by third parties.

Another relevant observation concerns the use of EVs, which have gained visibility and commercial interest in the past years. Currently, there are few studies [25], [41], [62], [80], [84] that have used or discussed the EV, and described how their applications behave in this type of vehicle, which is expected to become common in the future.

The quality of data creation is a challenge since different devices can create the same data, but with different levels of quality. Paefgen *et al.* [44] described an example that compares the use of sensors with different levels of quality to the same application. They evaluated the driving behavior using sensor data from a smartphone and its inertial sensors, and a commercial sensor unit [83] connected directly to the vehicle's OBD port. Different sensors from both groups had different impacts in their evaluation.

The key points we highlight in this stage are: (i) the costs to embed sensors or smart devices in a vehicle and evaluate the experiments reduce the capability of initiatives to investigate a new and unexplored field; (ii) the commercial value of vehicular data may not allow its availability to the community; (iii) vehicular data privacy is a challenge to its broad use; (iv) owners of vehicular data keep their propriety to advance their studies and financial opportunities; and (v) quality of data creation is a challenge since different devices can create the same data with different levels of quality.

B. Data Acquisition

The heterogeneity of data sources used in ITS leads to several difficulties in the data acquisition process. We can highlight the costs involved in installing sensors on the roads, environments and vehicles, which limit their data coverage. Another difficulty is to implement incentive mechanisms to obtain data from a group of individuals who are eligible to participate in an experiment. There are some initiatives to increase the data acquisition like the EnviroCar project [37], which is still limited to a geographic area, and others [56], [88], [114], [134] with a significant coverage.

In most cases, the reasons for having a restricted data coverage are cost to obtain the data, lack of volunteers, security and privacy issues, and corporate/government policies. In particular, security and privacy is a significant issue when developing ITS applications and should concern the society at large, since these datasets typically contain personal information. Thus, an important research issue is how to anonymize these datasets in such a way that different initiatives can use these data to advance the state of art of transportation systems, but preserving their security and privacy.

Another issue is the computational infrastructure to store and process large amounts of vehicular data. Transportation systems can generate vast quantities of heterogeneous data, which might outpace the development of adequate infrastructures and tools to support them, following, for instance, the data cycle of the VDS depicted in Figure 10. In this scenario, it is important to rationalize, organize and implement useful long-term infrastructures for sustaining current and future vehicular data needs, including offline and online (real-time) use.

Notice that the data scalability and privacy may reflect the data to be acquired. The limited scalability, in general, occurs whenever the data stays in a local device, which limits its use to a restricted number of participants/volunteers [56], [57], [90], [91]. If the collected data is available on the Web [61], [87], [111], [112], [113], this allows a broader adoption of participants/volunteers who may contribute to this dataset. Privacy issues may have an opposite effect in terms of user's adoption.

Even the case of Media as Vehicular Sensor (MVS) presents restrictions. The access to social media may be difficult or restricted, despite its minimal acquisition cost. For instance, Twitter allows users to collect its data through an API, but according to some restriction terms[13] that need to be followed such as the number of requisitions sent to the server, time interval between requisitions, among others. Moreover, there are other media sources that do not offer a structured way to collect their data, or just present a limited data coverage. These restrictions reduce the data to be analyzed, which may introduce more imprecision, users' bias, and spatiotemporal inconsistency.

Datasets related to VDS come from a variety of sources, as mentioned before, such as vehicles, roads, mobile devices and social media. Most of these data sources provide raw data typically including both spatial and temporal details, or at least one of them. Location, distance, direction and shape describe the spatial data characteristics, and duration and time the temporal aspect. Depending on the application, it might be more difficult to process and analyze data that has just one of these two aspects.

In principle, the combination of spatial and temporal aspects of one or more datasets has the potential to enrich a piece of data, which is a relevant topic of study in ITS. This process is not just the process of putting together different datasets according to their spatiotemporal aspects since there might be incomplete, imprecise and erroneous data, besides other issues. At the end, we may end up with a combined dataset that presents different levels of data quality, which can limit its use.

The key points we highlight at this stage are: (i) ITS uses heterogeneous data sources that need to be found and acquired; (ii) the limited coverage of available data constitutes a barrier to a deeper research in ITS; (iii) security and privacy are major issues when developing ITSs applications, since these data may contain personal information; (iv) the need for a computational infrastructure to store and process large amounts of vehicular data; (v) social media platforms impose restrictions in their collecting process (e.g., number of requisitions and data availability); (vi) the data enrichment process is promising and valuable for ITS applications, but brings new challenges that need to be overcome before using it.

C. Data Preparation

The data preparation is a critical stage in any study, since it is in this step that datasets are prepared to be used in different applications. It is at this stage that designers could consider to

[13]https://developer.twitter.com/en/docs

have "reliable datasets" that will have a strong impact on the final results.

Despite the relevance of this stage, just over half of the analyzed studies explicitly mention the data preparation, whereas the others do not clarify the steps to prepare the data for the next stage. One typical data preparation procedure is the reduction of variables, which aims to keep the most relevant features of the dataset [51], [52], [84]. After that, most of the data from the VDS include spatial or temporal aspects, and the necessity to filter them depends on the application goals, making the resulting dataset adequate to its use.

The second non-trivial procedure of data preparation is to perform its corrections, which almost all studies mention. In this step, the data aspects describe problems such as outliers, conflict, incompleteness, ambiguity, correlation and disparateness. These problems are more related to the data itself than to the methodologies used to combine them, mostly because the data collected from sensors are inherently imperfect. Based on these facts, the efforts to develop applications to an ITS usually depend on the role of each heterogeneous data to the application goal. Moreover, there is an inherent complexity in processing these data, which typically does not any standard. This may become a barrier to do research in ITS.

In the following, we describe some of the data problems commonly found in the VDS, and propose some solutions. A fine data granularity usually allows a more valuable information about the entities of interest. The data granularity is a concerning aspect of data fusion, especially when dealing with applications that use rough sets and neither fine-grained nor coarse-grained information is beneficial for the final process.

Vagueness occurs in datasets where attributes are not well defined. The loose definition of attributes allows subjective measures, i.e., "fast" or "slow". This issue commonly occurs in data sources like Questionnaire and Media from VDSource. The subjectivity of data present in social media, for instance, calls for strategies that allow its understanding. Using a Natural Language Processing (NLP) approach [108] and its algorithms like Term Frequency-Inverse Document Frequency (TF-IDF) [112], Spell correction and Stop-word filter [111], Latent Dirichlet Allocation (LDA) [113], and regular [110] expression, it is possible to reduce the noise and subjectivity of texts written by users. Fuzzy logic may also be used to remove the subjective aspect of these datasets.

Another issue in data preparation is the identification of outliers, i.e., incorrect data points that need to be removed from the dataset. This process is completely data dependent and different techniques can be used to perform this filtering process. If outliers are left in the dataset, they may undermine the final solution, leading to imprecise results. Some of the filtering techniques to address this problem are Kalman [88], [90] and Particle Filtering.

Incomplete data is, intuitively, data with missing parts. These missing parts may lead to incorrect conclusions and, thus, must be addressed. A possible strategy is to use probabilistic solutions whenever a data is missing. Ambiguity in datasets is a manifestation of its imprecision, and happens when two occurrences in the dataset are assumed to be precise and exact. However, they differ from each other.

There are other common methods to filter and correct the raw data. A Simple Moving Average (SMA) can be used to smooth out the effect of unwanted noise from the sensor data [56], [81], [91], for instance. Besides, a band-pass and low-pass filter may remove sensor noise [78], [81]. The GPS incomplete data may be treated using a simple linear interpolation [51], [91]. As a general way to prepare the raw data, we noticed the use of equations and thresholds (e.g., Max, Min, Mean, Median, Standard Deviation, Derivative, and Variance) to obtain particular results [48], [90], [108].

All data sources, especially sensors, have a confidence degree. Whenever this confidence is lower than 100%, data is considered uncertain. Solutions to this problem include statistical inference and belief functions. The VDS is inherently disparate since there are sensors that assess different aspects in different units and scales. Using large quantities of diverse data allow the extraction of contextual information unable to be captured by physical sensors.

In summary, an important challenge in this stage is to find the best algorithm/method to apply to the raw data, aiming to treat and prepare the dataset for the next step.

The key points we highlight at this stage are: (i) find the best way to fit and fix the data to be used in the proposed solutions; (ii) perform a variable reduction to keep the most relevant and descriptive features of the dataset; (iii) correct the dataset, by identifying outliers, conflict, incompleteness, ambiguity, correlation, and disparateness; (iv) apply data fusion techniques to also fit and fix the raw data; (v) use whenever possible standards to overcome the complexity of this problem domain and facilitate the research in ITS.

D. Data Processing

The data processing of VDS leads to various new descriptive data, giving vast possibilities of ITS applications as mentioned in Section V. In the data processing stage, the operation forms leverage new aspects from raw or treated data. Depending on the investigation aims, a set of methods (e.g., mathematical operations, algorithms, models) can be applied to the data to produce a high-level data, allowing the developing of new applications and services. Even considering the relevance of this stage to the whole data process, not all studies mentioned here made clear the description of the data processing stage.

Based on examples of those operations, we know that the derivative of the vehicle speed is the acceleration, and $\Delta Distance/\Delta Speed$ gives the displacement time and may describe the driver's reaction time. The air drag force calculated by the equation $cd \times \rho \times Speed^2 \times A$, where cd corresponds to a drag coefficient, ρ the density of air, and A the characteristic of the frontal area of the body, can describe the fuel consumption in some weather conditions, for instance.

The research in the ITS field involves multidisciplinary expertise once the dataset come from a variety of sources and each one is frequently used and maintained by specific groups. For instance, the weather data are supervised by meteorology institutes, although it can be used to alert risks on the road.

TABLE IX
MOST USED CLASSES OF MACHINE LEARNING ALGORITHMS BY THE ITS APPLICATIONS

Authors	Machine Learning Algorithms					
	Classification	Regression	Clustering	Dimensionality Reduction	Neural Network	Time Series
[104], [78], [70], [73] [51], [42], [46], [111] [43], [113], [50], [72] [52], [112], [48], [56] [40]	✓					
[66], [84], [51], [56]				✓		
[66], [69], [40], [51]		✓				
[43], [81], [72], [91]						✓
[69], [39], [72]			✓			
[49], [70]					✓	

Another data source that influences the traffic flow is provided by the department of transportation as a semaphore and speed limit. These data can be used to measure or identify the traffic flow. Furthermore, we can consider the weather data as a data layer to the whole transportation system. This means that each data point of other datasets present in the VDS might be associated with a weather data point (weather condition at that point). This can help to understand the traffic behavior from the point of view of weather conditions. Thus, a challenge here is to extract useful information from Intra-Vehicular Sensor (IVS) to perform some correlation with Extra-Vehicular Sensor (EVS), leading to personalized services for drivers in ITS.

In this scenario, data fusion becomes a tremendous challenge given the heterogeneity among the Vehicular Data Sources (VDSources), asynchronous sensor operation, sensor errors and sensor noise. Furthermore, the computational infrastructure and the spatiotemporal aspects contribute to the efforts to fuse heterogeneous data. Khaleghi *et al.* [135] conducted a comprehensive study of methodologies that aim to solve problems related to heterogeneous data fusion. They elaborated a taxonomy of data fusion aspects describing problems such as outliers, conflict, incompleteness, ambiguity, correlation and disparateness.

Rettore *et al.* [57] developed a methodology to recommend the best gears by fusing the speed data, engine RPM data and throttle position data, based on a mathematical function to achieve low fuel consumption and CO_2 emissions. Almost all reviewed studies, which developed applications such as driving behavior and road event detection, deal with, somehow, a data fusion technique [51], [52], [73] that integrates multiple data sources to produce a more useful information than the individual data. Some of them applied IVD fusion and others EVD fusion to achieve their goals. However, the joint treatment of both fusion strategies is scarcely explored, being an important research topic for future of ITSs.

Another common aspect related to this stage is the use of Machine Learning (ML) techniques in data processing. Almost half of the studies aim to detect the driving behavior or road event using a machine learning technique. Leveraging the ideas discussed in [136], [137], Table IX shows the classes of ML algorithms in those proposals.

Next, we highlight the methods/algorithms applied by them: Extreme Learning Machine (ELM) [52], Random Forest/Decision Trees [40], [51], [56], Support Vector Machines (SVMs) [42], [46], [51], [70], [73], [78], [104], [111], [113] to classifier pothole, turn, driver, and driving behaviour. Logistic Regression [40], [51], [66], [69] to predict the driver, drivers' risk, recognition of driving events. K-mean clustering [39], [69], [72], Dimensionality Reduction Algorithms like Principal Component Analysis (PCA) [51], [56], [66], [84], Viterbi and Baum-Welch algorithms [72], Artificial Neural Network (ANN) [49], [70], Adaboost [70], K-Nearest Neighbors (KNN) classifier [43], [113], Naïve Bayes (NB) method [48], [50], [112], [113], and, finally, Hidden Markov Models (HMM) to define different driver's behavior based on observations [104]. We also observed the use of algorithms to treat the temporal data aspects of VDS. The Dynamic Time Warping (DTW) algorithm aims to find an optimal alignment among signal vectors, allowing to detect and distinguish driving events, driver styles [43], [72], [81], [91].

The key points we highlight at this stage are: (i) find the best algorithms/methodologies for data processing is an important and hard-task to the proposed solutions. (ii) extract useful information from Intra-Vehicular Data (IVD) to correlate them with Extra-Vehicular Data (EVD), to allow personalized services. This will become one of the top trends for future ITSs; (iii) data fusion plays an essential task in data processing given the data heterogeneity among the Vehicular Data Sources (VDSources), and other aspects that need to be considered such as asynchronous sensor operation, sensor errors and sensor noise; (iv) machine learning (ML) techniques have a special role in data processing, mainly in classification and prediction tasks.

E. Data Use

In this stage of the data cycle, the obtained results are associated to the final application or a virtual sensor (see Section V). Moreover, both cases may require multiple data cycles to achieve their goals. In this case, the output of the last cycle becomes the input data for the next one. In the following, we give an example of such scenario: suppose roads are equipped with inductive loop devices that sense the vehicle speed. The data preparation stage makes the appropriate corrections to the data. In the next step, a set of thresholds are applied to signal exceeded speed limits to a particular vehicle, which is the input for an application of traffic fines. Using the same signal, we can have another cycle responsible for providing the data distribution of traffic fines based on particular periods. Finally, this distribution may generate another application to the authorities to improve their decision-making policies in that area.

In our analysis, we observed applications which aim to detect the driver's behavior, and road and traffic conditions. In those cases, there is a need to expand and test their applications in a real-time environment, to measure how individual drivers vary their driving style along the time. However, in general, those studies use a small sample data due to costs and the lack of volunteers, but, as frequently mentioned, they would like to have a significant number of drivers to obtain more reliable results.

Another interesting point is the identification of factors that influence the driving style. Most studies combined a small number of sensors to classify the individual driving style and, in some cases, related them to a crash risk. Some of the identified factors that affect the driving style are individual and sociocultural elements such as gender, age, driving experience, personality, cognitive style, group and organization values, and regional culture. This shows the need of more investigations, aiming to provide better and precise solutions.

In the design of an ITS, different pieces of information are normally used by various entities, in particular, drivers, who might receive lots of information. In some studies, interested in measuring the driver's workload and its negative impact on the driver's safety [138], [139], [140], drivers are stimulated with a variety of information during their trips. This is a relevant research issue to understand the workload that might put a driver at risk, which is clearly influenced not only by environmental factors, but individual ones as well. This adds another level of complexity since we cannot expect that all individual aspects will have the same weight every day. For instance, depending on the amount of sleeping hours in the previous night, a driver might become more angrier or not.

The key points we highlight at this stage are: (i) the adoption of an application depends on a variety of issues such as financial cost, utility for the user, interface, privacy and security issues, and performance; (ii) usually, the studies used a small sample data due to costs and lack of volunteers, reducing the capacity of generalizing their results; (iii) solutions should consider individual and sociocultural factors to provide better and more precise results; (iv) the driver's workload during a trip might impact negatively the driver's safety depending on different factors, which need to be further investigated when developing applications for ITSs.

VII. Conclusion

The development of new applications and services for the ITS environment depends on the availability and study of large amounts of data, which leads to the Vehicular Data Space (VDS).

In this paper, we survey recent studies describing services and applications for ITSs, but focused on the data used by them. We introduced the concept of VDS, which is used to describe the vehicular scenario from the data perspective. We proposed a taxonomy, according to the Vehicular Data Source (VDSource), discussed the different data sources currently used in ITSs. Furthermore, we discussed the relationship between *Costs* to develop and use each VDSource and its respective *Granularity* and *Scalability*. We also categorized the applications (Security, Eco-driving, Traffic Monitoring and Management, General Purpose, and Infotainment), noticing that 64% and 16% of them only used Intra-Vehicular Data (IVD) and Extra-Vehicular Data (EVD) to develop their applications, respectively, whereas 20% dealt with both groups. This clearly shows some interesting opportunities to explore the EVD and the fusion between IVD and EVD. Finally, we presented and discussed some challenges and open issues related to the main topics we observed based on the five stages of the VDS (*Data Creation*, *Data Acquisition*, *Data Preparation*, *Data Processing*, and *Data Use*).

We also discussed the use of heterogeneous datasets to provide accurate methods for ITS applications. Thus, data fusion techniques have the potential to improve the accuracy of those applications, when there are several related descriptors. Some typical sensors used to model and identify the driver's behavior are acceleration longitudinal/3-axis, GPS, turning, and vehicle speed. Also constitute an opportunity, the generation of CO_2 emissions and fuel consumption reports, based on the investigations that use Intra-Vehicular Sensor (IVS). These reports can be sent to authorities who will be better informed when taking their decisions.

Our comprehensive literature review also showed that most of the data available in the VDS are freely available for the public or partially accessible by them. It is also clear that novel ITS applications will benefit from multiple heterogeneous datasets. Of course, this does not mean that a single variable represents a less descriptive scenario. On the contrary, in some cases the longitudinal acceleration, for instance, can identify dangerous driving maneuvers in real time, being a good solution for insurance companies.

Considering the Vehicular Data Space (VDS), the main contributions of this work are: (i) the need of more investigations to recognize driving styles, relating them to individual and sociocultural factors; (ii) real driving observations need more spatiotemporal coverage; (iii) the need to expand and test applications in real-time environments; (iv) acceleration longitudinal/3-axis, GPS, turning, and vehicle speed are the most used sensor data to model driving behavior; (v) there is a complexity inherent in the processing of heterogeneous data since there is no standardization; (vi) heterogeneous data fusion is a fundamental challenge to leverage the ITS field.

References

[1] World Bank. (May 2017). *The World Bank*. [Online]. Available: http://www.worldbank.org/

[2] D. Schrank, B. Eisele, and T. Lomax, "TTI's 2012 urban mobility report," Texas A&M Transp. Inst., Texas A&M Univ. Syst., College Station, TX, USA, Rep., p. 4, 2012.

[3] D. Schrank, B. Eisele, T. Lomax, and J. Bak, "2015 urban mobility scorecard," Texas A&M Transp. Inst., Texas A&M Univ. Syst., College Station, TX, USA, Rep., p. 4, 2015.

[4] United States Environmental Protection Agency. (May 2017). *SmartWay*. [Online]. Available: https://www.epa.gov/smartway

[5] European Commission. (May 2017). *eSafety*. [Online]. Available: https://ec.europa.eu/

[6] U.S. Department of Transportation. (May 2017). *U.S. Department of Transportation*. [Online]. Available: https://www.transportation.gov/

[7] (May 2017). *European Transport Safety Council*. [Online]. Available: http://etsc.eu/

[8] U.S. Department of Transportation. (May 2017). *DSSS—UTMS Society of Japan*. [Online]. Available: http://www.utms.or.jp/english/system/dsss.html

[9] (May 2017). *Japan Transport Safety Board*. [Online]. Available: https://www.mlit.go.jp/jtsb/english.html

[10] Thyssenkrupp. (2017). *Urban-Hub: People Shaping Cities*. [Online]. Available: http://www.urban-hub.com/smart-mobility/

[11] ClickutilityTeam, Innovability, GL Events. (2017). *Smart Mobility Worlds*. [Online]. Available: http://www.smartmobilityworld.net/en/

[12] *Fundamentals of Automotive*, Texas Instrum., Dallas, TX, USA, 2018. [Online]. Available: http://tiautomotivefundamentals.com/

[13] "Connected car industry report—Part 2," Machina Res., Telefonica, Madrid, Spain, Rep., 2013.

[14] M. Faezipour, M. Nourani, A. Saeed, and S. Addepalli, "Progress and challenges in intelligent vehicle area networks," *Commun. ACM*, vol. 55, no. 2, pp. 90–100, 2012.

[15] K. Bengler *et al.*, "Three decades of driver assistance systems: Review and future perspectives," *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 4, pp. 6–22, Oct. 2014.

[16] F. Qu, F.-Y. Wang, and L. Yang, "Intelligent transportation spaces: Vehicles, traffic, communications, and beyond," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 136–142, Nov. 2010.

[17] O. K. Tonguz, H.-M. Tsai, T. Talty, A. Macdonald, and C. Saraydar, "RFID technology for intra-car communications: A new paradigm," in *Proc. IEEE Veh. Technol. Conf.*, Montreal, QC, Canada, Sep. 2006, pp. 1–6.

[18] O. K. Tonguz, H.-M. Tsai, C. Saraydar, T. Talty, and A. Macdonald, "Intra-car wireless sensor networks using RFID: Opportunities and challenges," in *Proc. Mobile Netw. Veh. Environ.*, 2007, pp. 43–48.

[19] M. Ahmed *et al.*, "Intra-vehicular wireless networks," in *Proc. IEEE Globecom Workshops*, Washington, DC, USA, Nov. 2007, pp. 1–9. [Online]. Available: http://ieeexplore.ieee.org/document/4437827/

[20] H.-M. Tsai *et al.*, "ZigBee-based intra-car wireless sensor networks: A case study," *IEEE Wireless Commun.*, vol. 14, no. 6, pp. 67–77, Dec. 2007.

[21] R. de Francisco, L. Huang, and G. Dolmans, "Coexistence of ZigBee wireless sensor networks and Bluetooth inside a vehicle," in *Proc. IEEE 20th Int. Symp. Pers. Indoor Mobile Radio Commun.*, Tokyo, Japan, Sep. 2009, pp. 2700–2704.

[22] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet Things J.*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[23] S. Reis, D. Pesch, B.-L. Wenning, and M. Kuhn, "Intra-vehicle wireless sensor network communication quality assessment via packet delivery ratio measurements," in *Mobile Networks and Management*, R. Agüero, Y. Zaki, B.-L. Wenning, A. Förster, and A. Timm-Giel, Eds. Cham, Switzerland: Springer Int., 2017, pp. 88–101.

[24] S. Tuohy *et al.*, "Intra-vehicle networks: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 534–545, Apr. 2015.

[25] J. Engelbrecht, M. J. Booysen, F. J. Bruwer, and G.-J. van Rooyen, "Survey of smartphone-based sensing in vehicles for intelligent transportation system applications," *IET Intell. Transp. Syst.*, vol. 9, no. 10, pp. 924–935, Dec. 2015.

[26] S. Abdelhamid, H. S. Hassanein, and G. Takahara, "Vehicle as a resource (VaaR)," *IEEE Netw.*, vol. 29, no. 1, pp. 12–17, Jan./Feb. 2015.

[27] H. Hartenstein and L. P. Laberteaux, "A tutorial survey on vehicular ad hoc networks," *IEEE Commun. Mag.*, vol. 46, no. 6, pp. 164–171, Jun. 2008.

[28] H. Hartenstein and K. Laberteaux, *VANET Vehicular Applications and Inter-Networking Technologies*, vol. 1. Torquay, U.K.: Wiley, 2009.

[29] G. Karagiannis *et al.*, "Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 4, pp. 584–616, 4th Quart., 2011.

[30] U. Lee and M. Gerla, "A survey of urban vehicular sensing platforms," *Comput. Netw.*, vol. 54, no. 4, pp. 527–544, 2010.

[31] J. Jeong and T. Oh, "Survey on protocols and applications for vehicular sensor networks," *Int. J. Distrib. Sensor Netw.*, vol. 12, no. 8, pp. 1–17, 2016.

[32] S.-T. Jeng and L. Chu, "Tracking heavy vehicles based on weigh-in-motion and inductive loop signature technologies," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 632–641, Apr. 2015.

[33] (May 2017). *Weather*. [Online]. Available: https://weather.com

[34] K. Merritt. (Aug. 2017). *Socrata*. [Online]. Available: https://socrata.com

[35] (Sep. 2006). *Twitter*. [Online]. Available: https://twitter.com

[36] (Sep. 2006). *Waze*. [Online]. Available: https://www.waze.com/

[37] A. Bröring *et al.*, "enviroCar: A citizen science platform for analyzing and mapping crowd-sourced car sensor data," *Trans. GIS*, vol. 19, no. 3, pp. 362–376, 2015. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84933049571&partnerID=tZOtx3y1

[38] M. Reininger, S. Miller, Y. Zhuang, and J. Cappos, "A first look at vehicle data collection via smartphone sensors," in *Proc. IEEE Sensors Appl. Symp. (SAS)*, Zadar, Croatia, Apr. 2015, pp. 1–6.

[39] M. Van Ly, S. Martin, and M. M. Trivedi, "Driver classification and driving style recognition using inertial sensors," in *Proc. IEEE Intell. Veh. Symp. (IV)*, Jun. 2013, pp. 1040–1045.

[40] C. D'Agostino, A. Saidi, G. Scouarnec, and L. Chen, "Learning-based driving events recognition and its application to digital roads," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2155–2166, Aug. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7052375/

[41] J. Carmona, F. García, D. Martín, A. L. Escalera, and J. M. Armingol, "Data fusion for driver behaviour analysis," *Sensors*, vol. 15, no. 10, pp. 25968–25991, 2015. [Online]. Available: http://www.mdpi.com/1424-8220/15/10/25968/

[42] O. Kumtepe, G. B. Akar, and E. Yuncu, "Driver aggressiveness detection via multisensory data fusion," *EURASIP J. Image Video Process.*, vol. 2016, no. 1, p. 5, 2016. [Online]. Available: http://jivp.eurasipjournals.com/content/2016/1/5

[43] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *Proc. IEEE Conf. Intell. Transp. Syst. (ITSC)*, 2011, pp. 1609–1615.

[44] J. Paefgen, F. Kehr, Y. Zhai, and F. Michahelles, "Driving behavior analysis with smartphones: Insights from a controlled field study," in *Proc. 11th Int. Conf. Mobile Ubiquitous Multimedia (MUM)*, 2012, Art. no. 36. [Online]. Available: http://doi.acm.org/10.1145/2406367.2406412

[45] N. AbuAli, "Advanced vehicular sensing of road artifacts and driver behavior," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2015, pp. 45–49. [Online]. Available: http://ieeexplore.ieee.org/document/7405452/

[46] C. Zhang *et al.*, "Driver classification based on driving behaviors," in *Proc. 21st Int. Conf. Intell. User Interfaces (IUI)*, 2016, pp. 80–84. [Online]. Available: http://doi.acm.org/10.1145/2856767.2856806

[47] R. K. Satzoda and M. M. Trivedi, "Drive analysis using vehicle dynamics and vision-based lane semantics," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 1, pp. 9–18, Feb. 2015. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6910285

[48] V. C. Magaña and M. Muñoz-Organero, "WATI: Warning of traffic incidents for fuel saving," *Mobile Inf. Syst.*, vol. 2016, Jan. 2016, Art. no. 3091516.

[49] J. E. Meseguer, C. T. Calafate, J. C. Cano, and P. Manzoni, "DrivingStyles: A smartphone application to assess driver behavior," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Split, Croatia, Jul. 2013, pp. 535–540. [Online]. Available: http://ieeexplore.ieee.org/document/6755001/

[50] J.-H. Hong, B. Margines, and A. K. Dey, "A smartphone-based sensing platform to model aggressive driving behaviors," in *Proc. 32nd Annu. ACM Conf. Human Factors Comput. Syst. (CHI)*, 2014, pp. 4047–4056. [Online]. Available: http://doi.acm.org/10.1145/2556288.2557321

[51] D. Hallac *et al.*, "Driver identification using automobile sensor data from a single turn," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Rio de Janeiro, Brazil, 2016, pp. 953–958.

[52] M. V. Martinez, J. Echanobe, and I. del Campo, "Driver identification and impostor detection based on driving behavior signals," in *Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC)*, Rio de Janeiro, Brazil, Nov. 2016, pp. 372–378. [Online]. Available: http://ieeexplore.ieee.org/document/7795582/

[53] H. Abut *et al.*, "Data collection with 'UYANIK': Too much pain; but gains are coming," in *Proc. Biennial DSP In-Vehicle Mobile Syst.*, Istanbul, Turkey, Jun. 2007.

[54] A. Riener and J. Reder, "Collective data sharing to improve on driving efficiency and safety," in *Proc. 6th Int. Conf. Autom. User Interfaces Interact. Veh. Appl. (AutomotiveUI)*, 2014, pp. 1–6. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-84910621204&partnerID=tZOtx3y1

[55] OpenXC. (Jan. 2012). *OpenXC*. [Online]. Available: http://openxcplatform.com/

[56] P. H. L. Rettore *et al.*, "Driver authentication in VANETs based on intra-vehicular sensor data," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Natal, Brazil, Jun. 2018, pp. 78–83.

[57] P. H. L. Rettore, A. B. Campolina, L. A. Villas, and A. A. F. Loureiro, "A method of eco-driving based on intra-vehicular sensor data," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Heraklion, Greece, Jul. 2017, pp. 1122–1127.

[58] M. Rutty, L. Matthews, J. Andrey, and T. D. Matto, "Eco-driver training within the City of Calgary's municipal fleet: Monitoring the impact," *Transp. Res. D Transp. Environ.*, vol. 24, pp. 44–51, Oct. 2013.

[59] CarChip. (Jun. 2013). *CarChip*. [Online]. Available: http://www.carchip.cc/

[60] M. Rutty, L. Matthews, D. Scott, and T. Del Matto, "Using vehicle monitoring technology and eco-driver training to reduce fuel use and emissions in tourism: A ski resort case study," *J. Sustain. Tourism*, vol. 22, no. 5, pp. 787–800, 2014. doi: 10.1080/09669582.2013.855221.

[61] K. Ayyildiz, F. Cavallaro, S. Nocera, and R. Willenbrock, "Reducing fuel consumption and carbon emissions through eco-drive training," *Transp. Res. F Psychol. Behav.*, vol. 46, pp. 96–110, Apr. 2017.

[62] C. Brace *et al.*, "Development and field trial of a driver assistance system to encourage eco-driving in light commercial vehicle fleets," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 796–805, Jun. 2013.

[63] J. Zhao, W. Li, J. Wang, and X. Ban, "Dynamic traffic signal timing optimization strategy incorporating various vehicle fuel consumption characteristics," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 3874–3887, Jun. 2016.

[64] PTV Group. (Aug. 1992). *VISSIM*. [Online]. Available: http://vision-traffic.ptvgroup.com/en-uk/home/

[65] R. Araújo, Â. Igreja, R. De Castro, and R. E. Araújo, "Driving coach: A smartphone application to evaluate driving efficient patterns," in *Proc. IEEE Intell. Veh. Symp.*, vol. 1. Alcala de Henares, Spain, 2012, pp. 1005–1010.

[66] C. Andrieu and G. S. Pierre, "Using statistical models to characterize eco-driving style with an aggregated indicator," in *Proc. IEEE Intell. Veh. Symp.*, Alcala de Henares, Spain, Jun. 2012, pp. 63–68. [Online]. Available: http://ieeexplore.ieee.org/document/6232197/

[67] J. F. R. Paefgen, "On the determination of accident risk exposure from vehicular sensor data—Methodological advancements and business implications for automobile insurance providers," Business dissertation, Univ. St. Gallen, St. Gallen, Switzerland, 2013.

[68] A. Mednis, A. Elsts, and L. Selavo, "Embedded solution for road condition monitoring using vehicular sensor networks," in *Proc. 6th Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, 2012, pp. 1–5.

[69] F. Guo and Y. Fang, "Individual driver risk assessment using naturalistic driving data," *Accident Anal. Prevention*, vol. 61, pp. 3–9, Dec. 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0001457512002382

[70] M. Elhenawy, A. Jahangiri, H. A. Rakha, and I. El-Shawarby, "Modeling driver stop/run behavior at the onset of a yellow indication considering driver run tendency and roadway surface conditions," *Accident Anal. Prevention*, vol. 83, pp. 90–100, Oct. 2015.

[71] Honda. (Nov. 2015). *Honda Sensing*. [Online]. Available: https://automobiles.honda.com/sensing

[72] F. Aloul, I. Zualkernan, R. Abu-Salma, H. Al-Ali, and M. Al-Merri, "iBump: Smartphone application to detect car accidents," *Comput. Elect. Eng.*, vol. 43, pp. 66–75, Apr. 2015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0045790615000713

[73] A. Fox, B. V. K. Kumar, J. Chen, and F. Bai, "Crowdsourcing undersampled vehicular sensor data for pothole detection," in *Proc. IEEE 12th Annu. Int. Conf. Sens. Commun. Netw. (SECON)*, Seattle, WA, USA, Jun. 2015, pp. 515–523. [Online]. Available: http://ieeexplore.ieee.org/document/7338353/

[74] Mechanical Simulation Corporation. (Jun. 2010). *CarSim*. [Online]. Available: https://www.carsim.com/

[75] J. Goncalves, J. S. V. Goncalves, R. J. F. Rossetti, and C. Olaverri-Monreal, "Smartphone sensor platform to study traffic conditions and assess driving performance," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Qingdao, China, Oct. 2014, pp. 2596–2601. [Online]. Available: http://ieeexplore.ieee.org/document/6958106/

[76] H. Han *et al.*, "SenSpeed: Sensing driving conditions to estimate vehicle speed in urban environments," in *Proc. IEEE INFOCOM*, 2014, pp. 727–735.

[77] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 16–55, May 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7926916/

[78] H. Chu *et al.*, "I am a smartphone and I know my user is driving," in *Proc. 6th Int. Conf. Commun. Syst. Netw. (COMSNETS)*, Bangalore, India, Jan. 2014, pp. 1–8.

[79] R. Vaiana *et al.*, "Driving behavior and traffic safety: An acceleration-based safety evaluation procedure for smartphones," *Mod. Appl. Sci.*, vol. 8, no. 1, p. 88, Jan. 2014. [Online]. Available: http://www.ccsenet.org/journal/index.php/mas/article/view/29037

[80] S. Kaplan, M. A. Guvensan, A. G. Yavuz, and Y. Karalurt, "Driver behavior analysis for safe driving: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3017–3032, Dec. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7225158/

[81] J. Engelbrecht, M. J. Booysen, and G.-J. Van Rooyen, "Recognition of driving manoeuvres using smartphone-based inertial and GPS measurement," in *Proc. 1st Int. Conf. Mobile Inf. Commun. Technol. Africa (UMICTA)*, Stellenbosch, South Africa, Dec. 2014, pp. 88–92.

[82] M. Fazeen, B. Gozick, R. Dantu, M. Bhukhiya, and M. C. González, "Safe driving using mobile phones," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1462–1468, Sep. 2012.

[83] Scope Technology. (Jul. 1999). *Scope Technology*. [Online]. Available: http://www.scopetechnology.com/

[84] G. Castignani, T. Derrmann, R. Frank, and T. Engel, "Driver behavior profiling using smartphones: A low-cost platform for driver monitoring," *IEEE Intell. Transp. Syst. Mag.*, vol. 7, no. 1, pp. 91–102, Jan. 2015.

[85] Q. Yuan, Z. Liu, J. Li, S. Yang, and F. Yang, "An adaptive and compressive data gathering scheme in vehicular sensor networks," in *Proc. Int. Conf. Parallel Distrib. Syst. (ICPADS)*, vol. 2016. Melbourne, VIC, Australia, 2016, pp. 207–215.

[86] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, "Crowd sensing of traffic anomalies based on human mobility and social media," in *Proc. 21st ACM SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst. (SIGSPATIAL)*, Orlando, FL, USA, 2013, pp. 334–343. [Online]. Available: http://doi.acm.org/10.1145/2525314.2525343

[87] S. Wang *et al.*, "Computing urban traffic congestions by incorporating sparse GPS probe data and social media data," *ACM Trans. Inf. Syst.*, vol. 35, no. 4, pp. 1–30, Jul. 2017. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3112649.3057281

[88] L. M. Bergasa, D. Almerá, J. Almazán, J. J. Yebes, and R. Arroyo, "DriveSafe: An app for alerting inattentive drivers and scoring driving behaviors," in *Proc. IEEE Intell. Veh. Symp.*, Dearborn, MI, USA, Jun. 2014, pp. 240–245.

[89] AXA. (Jul. 2013). *AXA Drive*. [Online]. Available: https://www.axa.com

[90] C. Ma *et al.*, "DrivingSense: Dangerous driving behavior identification based on smartphone autocalibration," *Mobile Inf. Syst.*, vol. 2017, Mar. 2017, Art. no. 9075653.

[91] C. Saiprasert, T. Pholprasit, and S. Thajchayapong, "Detection of driving events using sensory data on smartphone," *Int. J. Intell. Transp. Syst. Res.*, vol. 15, no. 1, pp. 17–28, 2017.

[92] Z. Wang, M. Lu, X. Yuan, J. Zhang, and H. Van De Wetering, "Visual traffic jam analysis based on trajectory data," *IEEE Trans. Vis. Comput. Graphics*, vol. 19, no. 12, pp. 2159–2168, Dec. 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6634174/

[93] D. Parker, J. T. Reason, A. S. R. Manstead, and S. G. Stradling, "Driving errors, driving violations and accident involvement," *Ergonomics*, vol. 38, no. 5, pp. 1036–1048, 1995.

[94] A. I. Glendon *et al.*, "Reliability of the driving behaviour inventory," *Ergonomics*, vol. 36, no. 6, pp. 719–726, 1993.

[95] D. J. French, R. J. West, J. Elander, and J. M. Wilding, "Decision-making style, driving style, and self-reported involvement in road traffic accidents," *Ergonomics*, vol. 36, no. 6, pp. 627–644, 1993.

[96] H. A. Deery and A. W. Love, "The driving expectancy questionnaire: Development, psychometric assessment and predictive utility among young drink-drivers," *J. Stud. Alcohol Drugs*, vol. 57, no. 2, pp. 193–202, 1996.

[97] V. Beanland, M. Fitzharris, K. L. Young, and M. G. Lenné, "Driver inattention and driver distraction in serious casualty crashes: Data from the Australian national crash in-depth study," *Accid. Anal. Prevent.*, vol. 54, pp. 99–107, May 2013. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S000145751300047X

[98] F. M. Poó and R. D. Ledesma, "A study on the relationship between personality and driving styles," *Traffic Injury Prevent.*, vol. 14, no. 4, pp. 346–352, Mar. 2013.

[99] M. Zuckerman, "Zuckerman–Kuhlman personality questionnaire (ZKPQ): An alternative five-factorial model," in *Big Five Assessment*. Seattle, WA, USA: Hogrefe & Huber, 2002, pp. 377–396.

[100] H. H. van Huysduynen, J. Terken, J.-B. Martens, and B. Eggen, "Measuring driving styles: A validation of the multidimensional driving style inventory," in *Proc. ACM 7th Int. Conf. Autom. User Interfaces Interact. Veh. Appl. (AutomotiveUI)*, Nottingham, U.K., 2015, pp. 257–264. [Online]. Available: http://doi.acm.org/10.1145/2799250.2799266

[101] O. Taubman-Ben-Ari, M. Mikulincer, and O. Gillath, "The multidimensional driving style inventory–scale construct and validation," *Accid. Anal. Prevent.*, vol. 36, no. 3, pp. 323–332, 2004.

[102] F. Sagberg, Selpi, G. F. B. Piccinini, and J. Engström, "A review of research on driving styles and road safety," *Human Factors*, vol. 57, no. 7, pp. 1248–1275, 2015. [Online]. Available: https://doi.org/10.1177/0018720815591313

[103] D. M. Truxillo, J. Macarthur, L. B. Hammer, and T. N. Bauer, "Evaluation of a supervisor training program for ODOT's EcoDrive program," Transp. Res. Educ. Center, Portland, OR, USA, Rep. NITC-RR-781, 2016.

[104] G. S. Aoude, V. R. Desaraju, L. H. Stephens, and J. P. How, "Behavior classification algorithms at intersections and validation using naturalistic data," in *Proc. IEEE Intell. Veh. Symp. (IV)*, Baden-Baden, Germany, 2011, pp. 601–606.

[105] Open Weather Map. (May 2017). *OpenWeatherMap*. [Online]. Available: https://openweathermap.org/api

[106] B. P. Santos, P. H. L. Rettore, H. S. Ramos, L. F. M. Vieira, and A. A. F. Loureiro, "Enriching traffic information with a spatiotemporal model based on social media," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Natal, Brazil, Jun. 2018, pp. 00464–00469.

[107] P. H. L. Rettore, I. Araujo, J. G. M. de Menezes, L. Villas, and A. A. F. Loureiro, "Serviço de detecção e enriquecimento de eventos rodoviários baseado em fusão de dados heterogêneos para vanets," in *Proc. SBRC*, Gramado, Brazil, May 2019. [Online]. Available: http://sbrc2019.sbc.org.br/sessoes-tecnicas/

[108] Y. Gu, Z. S. Qian, and F. Chen, "From Twitter to detector: Real-time traffic incident detection using social media data," *Transp. Res. C Emerg. Technol.*, vol. 67, pp. 321–342, Jun. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0968090X16000644 http://linkinghub.elsevier.com/retrieve/pii/S0968090X16000644

[109] I. Septiana, Y. Setiowati, and A. Fariza, "Road condition monitoring application based on social media with text mining system: Case study: East Java," in *Proc. IEEE Int. Electron. Symp. (IES)*, Denpasar, Indonesia, Sep. 2016, pp. 148–153. [Online]. Available: http://ieeexplore.ieee.org/document/7860992/

[110] H. Shekhar, S. Setty, and U. Mudenagudi, "Vehicular traffic analysis from social media data," in *Proc. IEEE Int. Conf. Adv. Comput. Commun. Informat. (ICACCI)*, Jaipur, India, Sep. 2016, pp. 1628–1634. [Online]. Available: http://ieeexplore.ieee.org/document/7732281/

[111] M. Sinha, P. Varma, and T. Mukherjee, "Web and social media analytics towards enhancing urban transportations: A case for Bangalore," in *Proc. ACM 2nd Int. Workshop Netw. Data Anal. (NDA)*, Chicago, IL, USA, 2017, pp. 1–7. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3068943.3068950

[112] A. Kurkcu, F. Zuo, J. Gao, E. F. Morgul, and K. Ozbay, "Crowdsourcing incident information for disaster response using Twitter," in *Proc. 96th Annu. Meeting Transp. Res. Board*, Washington, DC, USA, Jan. 2017.

[113] R. Y. K. Lau, "Toward a social sensor based framework for intelligent transportation," in *Proc. IEEE 18th Int. Symp. World Wireless Mobile Multimedia Netw. (WoWMoM)*, Macau, China, Jun. 2017, pp. 1–6. [Online]. Available: http://ieeexplore.ieee.org/document/7974354/

[114] P. Angkititrakul *et al.*, "UTDrive: The smart vehicle project," in *In-Vehicle Corpus and Signal Processing for Driver Behavior*, K. Takeda, H. Erdogan, J. H. L. Hansen, and H. Abut, Eds. Boston, MA, USA: Springer, 2009, pp. 55–67. [Online]. Available: https://doi.org/10.1007/978-0-387-79582-9_5

[115] J. Paefgen *et al.*, "Telematics strategy for automobile insurers: Whitepaper," Working Paper, Inst. Technol. Manag., Transf. Center Technol. Manag., St. Gallen, Switzerland, 2013. [Online]. Available: https://www.alexandria.unisg.ch/223402/

[116] *Modeling the Relation Between Driving Behavior and Fuel Consumption*, CGI, Montreal, QC, Canada, p. 10, 2014.

[117] MirrorLink. (May 2017). *MirrorLink—Car Connectivity Consortium*. [Online]. Available: http://www.mirrorlink.com/

[118] Ford. (Jul. 2010). *AppLink*. [Online]. Available: https://developer.ford.com/pages/applink

[119] Magister54. (Jul. 2015). *OpenGauge*. [Online]. Available: https://github.com/Magister54/opengauge

[120] *Highway Statistics 2015*, Federal Highway Admin., Washington, DC, USA, 2016.

[121] P. Schroeder, M. Meyers, and L. Kostyniuk, "National survey on distracted driving attitudes and behaviors–2012," Nat. Highway Traffic Safety Admin., Washington, DC, USA, Rep., 2013.

[122] W. Wang, J. Xi, and H. Chen, "Modeling and recognizing driver behavior based on driving data: A survey," *Math. Problems Eng.*, vol. 2014, p. 20, Feb. 2014.

[123] StateFarm. (Jul. 2017). *Drive Safe and Save*. [Online]. Available: https://www.statefarm.com/insurance/auto

[124] Crash Detech. (Jun. 2017). *Crash Detech*. [Online]. Available: https://www.crashdetech.com

[125] A. L. L. Aquino, T. S. G. Cavalcante, E. S. Almeida, A. C. Frery, and O. A. Rosso, "Characterization of vehicle behavior with information theory," *Eur. Phys. J. B*, vol. 88, no. 10, p. 257, Oct. 2015. [Online]. Available: https://doi.org/10.1140/epjb/e2015-60384-x

[126] M. J. Silva *et al.*, "Study about vehicles velocities using time causal information theory quantifiers," *Ad Hoc Netw.*, vol. 89, pp. 22–34, Jun. 2019.

[127] GM. (May 2011). *OnStar*. [Online]. Available: https://www.onstar.com/us/en/home.html

[128] Audi. (May 2014). *Audi Connect*. [Online]. Available: https://www.audiusa.com/help/audi-connect

[129] SDL Consortium, Inc. (May 2017). *Smart Device Link—SDL*. [Online]. Available: https://www.smartdevicelink.com

[130] H. T. Cheng, H. Shan, and W. Zhuang, "Infotainment and road safety service support in vehicular networking: From a communication perspective," *Mech. Syst. Signal Process.*, vol. 25, no. 6, pp. 2020–2038, 2011.

[131] Apple. (May 2014). *Car Play—Apple*. [Online]. Available: https://www.apple.com/ios/carplay/

[132] Toyota. (May 2015). *Toyota Touch 2*. [Online]. Available: https://www.toyota-europe.com/world-of-toyota/articles-news-events/2016/toyota-touch-2

[133] BMW. (May 2014). *BMW ConnectedDrive*. [Online]. Available: http://www.bmwusa.com/standard/content/innovations/bmwconnecteddrive/connecteddrive.aspx

[134] R. Rawassizadeh and D. Kotz, "Datasets for mobile, wearable and IoT research," *GetMobile Mobile Comput. Commun.*, vol. 20, no. 4, pp. 5–7, 2017.

[135] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Inf. Fusion*, vol. 14, no. 1, pp. 28–44, Jan. 2013.

[136] A. Ferdowsi, U. Challita, and W. Saad, "Deep learning for reliable mobile edge analytics in intelligent transportation systems: An overview," *IEEE Veh. Technol. Mag.*, vol. 14, pp. 62–70, 2017.

[137] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," *arXiv preprint arXiv:1710.02913*, 2017.

[138] J.-M. Girard, K. Younsi, N. Tricot, and J. C. Popieul, "When does the driver workload reaches its limits?" in *Proc. IEEE Intell. Transp. Syst. Conf.*, Toronto, ON, Canada, Sep. 2006, pp. 578–583.

[139] L. Angell *et al.*, "Driver workload metrics project: Task 2 final report," U.S. Dept. Transp., Washington, DC, USA, p. 460, 2006.

[140] C. Wu and Y. Liu, "Queuing network modeling of driver workload and performance," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 3, pp. 528–537, Sep. 2007.