

TOWARD FLEXIBLE DATA COLLECTION OF DRIVING BEHAVIOUR

M. AMEKSA¹, H. MOUSANNIF¹, H. AL MOATASSIME², Z. ELAMRANI ABOU ELASSAD¹

¹ LISI Laboratory, Computer Science Department, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco - mohammed.ameksa@ced.uca.ma, mousannif@uca.ac.ma, z.elamrani@uca.ma

² OSER Research Team, Computer Science Department, Faculty of Sciences and Technologies, Cadi Ayyad University, Marrakech, Morocco - hassan.al.moatassime@gmail.com

KEY WORDS: Data Collection, Driving Behavior, DVE's Data, Driving Dimensions.

ABSTRACT:

Recently, driving behavior has been the focus of several researchers and scientists, they are attempting to identify and analyze driving behavior using different sources of data. The purpose of this research is to investigate data acquisition methods and tools related to driving behavior, in addition to the type of data acquired. Using a systematic literature review strategy, this study identified tools and techniques used to collect data related to driving behavior among 120 selected studies from 2010 to 2020 in several literature resources. It then measured the percentages of the most commonly used methods, as well as the type of data collected. In-vehicle and IoT sensors was found to play the greatest role in data collection in approximately 67% of the documents selected studies; And concerning the type of data acquired, those relating to the vehicle are the most widely collected. Thus, this study definitively answers the question regarding the different data sources and data types used among researches. However, further studies are needed to give more attention to the driver's data and also to investigate the data from the three dimensions of driving (driver, vehicle, and environment) together as an integrated and interconnected system.

1. INTRODUCTION

Recent research indicates that driver error contributes to up to 75% of all roadway crashes (Stanton and Salmon, 2009). Literally, human factors contribute in the manifestation of 95% of all accidents, according to study of 2041 traffic accidents conducted by (Sabey and Taylor, 1980). Reducing those huge numbers and save people's life become necessity, for that reason and in order to improve safety, security and comfort of the driver and other road's users, many studies were dealing with the topic of driving behavior (DB) using different approaches and techniques.

The common element in these studies, is represented by source of data according to (Elamrani Abou Elassad and Mousannif, 2019). In fact, the majority of researches at the field of DB are using one of those three types of studies, Naturalistic Driving Studies (NDS), Field Driving Studies (FDS) or Simulator Driving Studies (SDS) to collect data (Yang et al., 2018a). Which clearly allow us to realize the importance of the data acquisition process to analyze driving behavior.

Moreover, (Andria et al., 2015) considered that the data acquisition in automotive environments is widely used in everyday applications. Actually, the recent computerizations of cars, together with the development of sensor technologies and car communication devices have transformed the cars into wealthy sources of information on the driver, the vehicle and environment (Bouhoute et al., 2019). In addition, the remarkable advancement of Internet of vehicles (IoV) technologies and big data technologies in recent years have offered new solutions to improve traffic safety and efficiency (Cen et al., 2017).

While these precursor works offer helpful insights into DB evaluation from a data-based perspective, it is crucial to note that through data-collection examination for DB analysis is

quite limited; to the authors' knowledge minimal work has been directed to the investigation of the harnessed data characteristics in this domain. Therefore, this paper aims to present a short survey that reveals methods of data collection process and tools linked to Driving Behavior, in which we present the most techniques and measures used to collect and gather useful information to analyze driving behavior. Another aspect has also been covered in this work which concerns the three driver's dimensions data.

The rest of the paper is organized as follows: Section II presents the methodology adopted to select some related existing works to data collection of driving behavior, and the process of extracting and synthesizing the data. In Section III, results obtained about techniques and technologies used to collect data, then the three dimensions of driving behavior related data. Finally, Section IV concludes the paper.

2. METHODOLOGY

In order to identify, analyze and interpret all available evidence related to "data collection in the area of driving behavior", we planned, conducted and reported the review by following the systematic literature review SLR process (often referred to as a systematic review) suggested by (Kitchenham and Charters, 2007).

This process aims to present a fair evaluation of the topic mentioned above using a trustworthy, rigorous, and auditable methodology.

Thus, this study adopted the three famous phases of SLR which are: planning the review, conducting the review and reporting the review (i.e. Figure 2).

Among the mandatory stages in the SLR is to confirm and to identify the need for such a review. In fact, defining the

research questions and producing a review protocol are the most important pre-review activities.

The first two phases are described by (Wen et al., 2012) through their development of the review protocol that mainly includes six stages (Figure 1): research questions definition, search strategy design, study selection criteria and procedures, quality

e) Use the Boolean AND to link the major terms from population, intervention, and outcome.

The result of analyzing RQs of topic “Toward flexible data collection of driving behavior” mentioned above brought us to

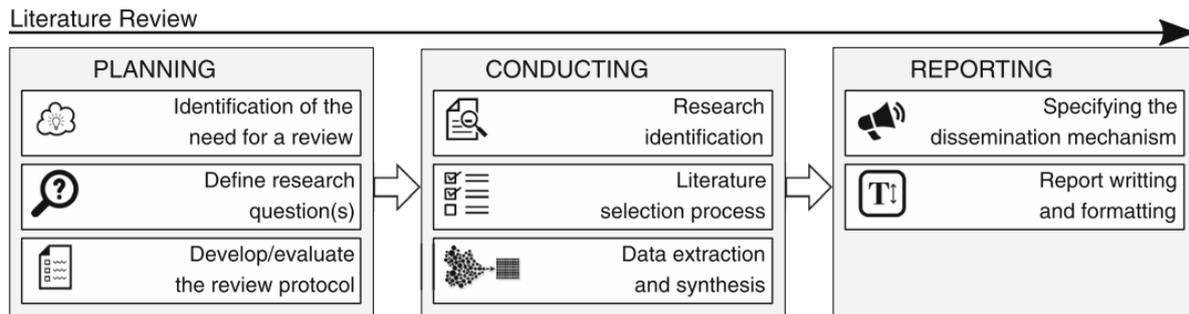


Figure 2. The stages of Systematic Literature Review

assessment, data extraction and data synthesis. The figure illustrates the whole process followed on this study. The first stage in this process involves raising a set of research questions (RQs) based on the main objective of the study.

2.1 Research questions

The aim of this paper is to summarize and describe the majority of techniques of data collection in the area of driving behavior including all its dimensions. Towards this aim some RQs were addressed. The table (Table 1) illustrate them:

ID	Research questions
RQ1	Which techniques are used to collect data in the area of driving behavior?
RQ2	which is/are the most important element (s) in driving behavior? is it drivers, vehicle, or environment's data??

Table 1. Research questions

2.2 Search strategy

Once the RQs have been identified, a research strategy must be followed it. It consists of selecting the search key terms (keywords), resources (libraries or others with relevant experience) and search process.

2.3 Search terms

The search terms used in this paper were constructed using the following strategy (Wen et al., 2012), (Kitchenham et al., 2007):

- Derive major terms from the questions;
- Identify alternative spellings and synonyms for major terms;
- Check the keywords in any relevant papers we already have;
- Use the Boolean OR to incorporate alternative spellings and synonyms;

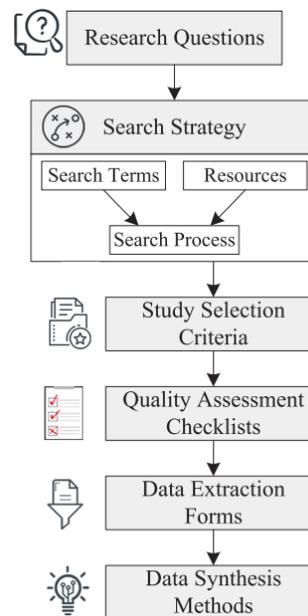


Figure 1 Stages of review protocol

extract the following keywords:

Data collection - Driving behavior

After that, we tried to find new words, synonyms and alternatives spellings of the keywords already found and the results are:

- Data collection: acquisition, assembling,
- Driving behavior: driving style, driving pattern, driving profile.

Once we identified the most keywords and their synonyms, we adopt the basic rule to establish the search string: for each separated word, we found its synonyms and concatenated them with the OR connector. After the definition of the groups of words with their synonyms, we concatenated them with AND to end the string. And search string extracted from are:

(“Data collection” OR “data acquisition”) AND (“driving behavior” OR “driving style” OR “driving pattern” OR “driving profile”)

2.4 Resources

In this study, we used four electronic databases as the literature resources to search for primary studies (IEEE Xplore, ScienceDirect, Web of Science and Google Scholar). Since the search engines of different databases use different syntax of search strings, our search string constructed previously was adjusted to accommodate different databases and used to search for journal papers in those electronic databases published between 2010 and 2020.

2.5 Search process

This research has been conducted on the four electronic databases separately, then export the CSV file of the returned papers and gather the results together to form a set of candidate papers (Figure 3). A script of python was applied to this set of papers to generate word cloud of the title and the abstract of each article. This script is free available on GitHub¹.

Then, the set of articles selected has been scanned so as to remove duplicated documents. Some reading strategies has been used and described on next subsection 'study selection criteria' to identify 120 relevant articles which were then used for data extraction and data synthesis.

2.6 Study selection criteria

Search criteria for a first stage resulted in 1224 candidate papers (see Figure 3). Due to the fact that many of the candidate documents do not provide any useful information to answer the research questions raised by this paper, further filtering is needed to identify the relevant papers. Knowing that both the title and the abstract are generally written correctly, accurately, carefully, and meticulously, i.e. they confirm whether the document is strongly pertinent for the mean topic of the study or not. Moreover, the 'word cloud' technique reveals the essential from

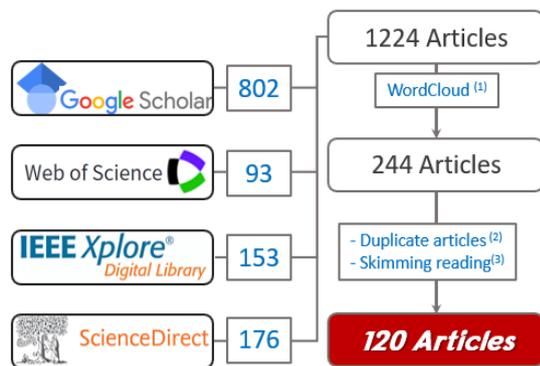


Figure 3. Process of selecting articles

- (1) Visual representation of the most used keywords
- (2) Remove duplicated articles using Mendeley software.
- (3) This strategy was applied on the title and abstract of each article.

an extract of text, fast and engaging. It was applied to the title and the abstract of every documents, it is used to represent the words that compose the title and the abstract in different sizes according to the frequency of their use, as illustrated in (Figure 4).

Analyzing the results obtained and keeping those in which the following words, 'data, behavior, driving, collecting and acquisition'. If one of the above-mentioned terms appeared widely and broadly, we select the article. As a result, we have selected 244 articles.

We have used the Skimming² and scanning³ reading's techniques for the purpose of removing the duplicated articles and to get a general overview of the relevant article. During this stage, we try to preserve scientific documents that contain valuable insights about the data collection, thus permitting us to highlight 120 relevant articles.



Figure 4. Example of word cloud generated from tow document

2.7 Study quality assessment

On the one hand, the quality assessment QA of the selected studies is initially used as the basis for weighting the quantitative data extracted in the meta-analysis according to (Julian PT Higgins, 2009). And since we are interested in this first work by the percentage of data sources used and the percentage of driving's dimension data on the other hand, we do not specify a dedicated QA to this paper. Instead, we just verified whether the articles involved provide relevant information regarding of all these aspects.

2.8 Data extraction forms:

This subsection aims to clarify the process of extraction the data followed in this paper. We exploited the selected studies to collect the data that contribute to addressing the research questions concerned in this work. In fact, the data extraction process is designed to answer the following questions:

- what is the data acquisition tool used to acquire the data?
- Which of the three dimensions of driving is covered? driver, vehicle, or environment?

While trying to find answers to these questions, some data could not be extracted directly from the selected studies. Nevertheless, we were able to obtain them indirectly by processing the available data in an appropriate form. For example, there are some studies that use databases offered by other previous works, in this case we try to see sources of data in the original work if available, otherwise we conclude based on the rest of the article. (Figure 5) illustrates some extracted data. As shown, the figure composed of three headings. These rubrics

² reading a text quickly to get a general idea of meaning.

³ reading in order to find specific information, e.g. figures or names.

¹ <https://github.com/AMEKSA/ResearchMethodology>

are in a way a reformulation of the previous questions. The data extraction process consists of giving "1" or "0" (green icon or red icon) according to the presence or absence respectively of each item in the article, the comprehensive list of the relevant articles selected to this paper and extraction results are given in the appendix.

2.9 Data synthesis methods

Authors	Source of data						Dimension's data		
	In-vehicle sensors	Self-report questionnaires	Other sensors	smartphone	Traffic surv	Dataset	Driver	Vehicle	roadway environment
(Zhou <i>et al.</i> , 2011)	✗	✗	✗	✗	✓	✗	✗	✓	✗
(Abbas <i>et al.</i> , 2011)	✓	✗	✓	✗	✗	✗	✓	✓	✗
(Takeda <i>et al.</i> , 2011)	✓	✗	✓	✗	✗	✗	✓	✓	✓

Figure 5. Example of extracted data

Data synthesis aims to gather all previous results, interpret results, shed light on the interests of most researchers and reveal some future areas of research. Actually, the purpose of data synthesis is to aggregate evidence from the selected studies for answering the research questions. Therefore, by summing up the scores obtained through results of data extraction process and using some visualization tools, including pie chart to present the percentages pertaining to the source of data used and DVE's data among all the selected articles, we can move to the next section which will be dedicated to results and discussion.

3. RESULTS AND DISCUSSION

This section presents and discusses the findings of this short review. First, we introduce the data collection topic and the statistics of most commonly methods used to gather information related to driving behavior. Then, we present the instruments and measurements techniques used to collect data according to the selected studies one by one in the separate subsections. Second, we reported statistics of researchers' attention to driver's dimensions.

3.1 Data collection

This section aims to shed light upon the data collection process and the most techniques used in the literature research related with the field of driving behavior. Before citing methods of data collection, it appears necessarily to define what is the data acquisition first?

According to Cambridge Dictionary (Cambridge), data collection activity means collecting information that can be used to find out about a particular subject. This activity enables a person or organization to understand the relevance topic, answer its linked questions, evaluate outcomes, and make predictions about future probabilities and trends. So, in order to understand driving behavior and a major factor for road traffic safety, assembling and gathering its associated data is a mandatory stage. Nevertheless, the varieties on the sources of data cause a difference in understanding driving behavior among researchers. In fact, the studies on driving behavior assessment have not settled on a common framework due to this diversity (Zhu *et al.*, 2017).

As mentioned above, it is assumed that the rest of this section will describe techniques used in data collection. Some driving style-related studies used self-report and driving behavior questionnaires to collect information, other several studies have

taken advantage of new technologies and benefit from the incredible development of automotive sensors such as In-Vehicle Data Recorders, smartphones, IoT sensors and traffic surveillance technologies to sense and collect contributions attributes of DB.

In short, according to (Carvalho *et al.*, 2017), the data collected from the action of driving can be carried out by several kinds of sensors, from those of a general kind in smartphones to

dedicated devices such as monitoring cameras, telematics boxes⁴ and OBD⁵ (On-Board Diagnostic) adapter. (Figure 6) illustrates the statistics and the percentages of techniques and methods used to collect data for driving behavior study according to scientific researches selected for this work.

A brief description of these techniques is detailed in (Hata! Başvuru kaynağı bulunamadı.) as shown below.

As can be seen in (Figure 6), the most dominant data source for driving behavior among the 120 articles used in this paper is the

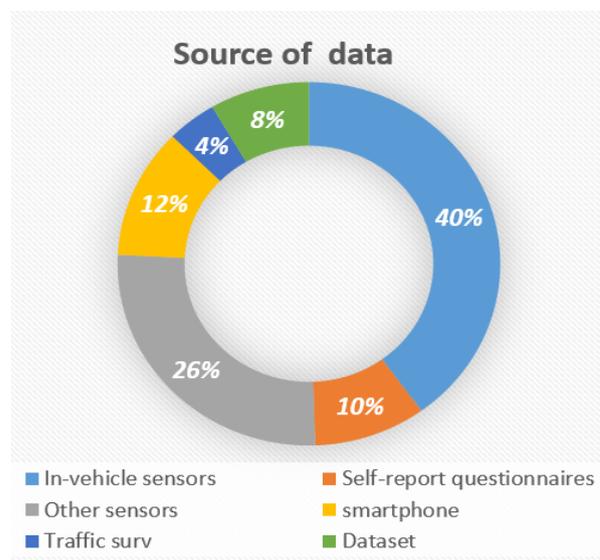


Figure 6. Collecting data sources used on literature studies

integrated sensors "In-vehicle sensors", at the rate of 40%. Then IoT device and other sensors with a percentage of 26%. The smartphone has also demonstrated a strong capacity of data collection at the rate of 12%, followed by the use of Self-report technique by 10% and other databases with 8%. Finally, the use of the traffic surveillance's tools by 4%.

⁴ A telematics box or black box is a measurement probe installed inside the vehicle. It may be equipped with its own sensors or be connected to the vehicle's internal sensors via the CAN-bus.

⁵ OBD is a system that enables current vehicles to carry out a self-diagnosis and provide real time data (e.g., speed) via a standard communications port.

Therefore, the type of sensors embedded in the vehicle remains the best source to gather data according to the literature used for this study.

Source	Description
In-vehicle sensors	Devices which can transform physical quantities such as pressure or acceleration into output signals (usually electrical). And always embedded on the vehicle.
Self-report questionnaires	A research instrument consists of a series of questions to gather information and data about the driver.
Other sensors	sensors that are not integrated into the car, including sensors of IoT like Arduino and Raspberry Pi, ...
smartphone	high-resolution and high-speed (CMOS) image sensor, global positioning system (GPS) sensor, accelerometer, gyroscope, ambient light sensor, and microphone,
Traffic surveillances	Observation from a distance, using some techniques such as closed-circuit television (CCTV), or interception of electronically transmitted information such as internet traffic.
Dataset	International driving-dataset projects.

Table 2. Description of measurement techniques used to collect data

3.1.1 In-vehicle and other sensors: Driving behavior related data can be acquired usually by on-board devices. In Vehicle Data Recorders (IVDRs) are one of the tools widely used for on-board data collection. They are devices installed on vehicles that monitor and record continuously the vehicle parameters (Bouhoute et al., 2019). In reality, car sensors can produce about 1.3 gigabytes of data every hour and an estimated 312 million gigabytes every year for 4 hours of daily driving according to IBM (Kimberly Madia, 2014), which provides a valuable opportunity for researchers in the area of Driving Behavior. Based on those sensors, (Boquete et al., 2010), (Xie et al., 2019), among other researchers, present a various platforms as the acquisition system. However, the major limitation of this approach is the availability of OBU. It is likely that such an advanced technology is only available to a biased subset of vehicles.

Furthermore, the need of data from the driver's interaction with the vehicle requires more sensors to be added to the vehicle. For this reason, several demonstration tools have been developed to access the available telemetry data, (Ding et al., 2019) Used electroencephalography (EEG) and steering behavior in a simulated driving experiment to test the correlation between some patterns of driving behavior, cognitive states and personalities. Since driving is a social act which human factor has the most important role in it. Some researchers are interested in the study of the human contribution. (Yang et al., 2018b) used an electrode cap connected to Curry 7 software to collect EEG signals.

3.1.2 Self-report and questionnaires: Studies in transportation psychology have traditionally employed self-report measures to examine personality, motivations, cognitions, and perceptions on the one hand, and driving behavior, driving styles and skills and involvement in traffic violations and crashes on the other. Nevertheless, the usefulness and validity of such instruments is often questioned particularly when the aim is to capture risky driving behavior (Boufous et al., 2010). Generally, there is widespread use of self-report measures of driving behavior in the traffic psychology literature. Moreover, Most prevailing studies have used subjective questionnaire data and objective driving data to classify driving behavior whereas few studies have used physiological signals such as electroencephalography (EEG) to gather data (Yang et al., 2018b). One of the studies adopted self-report technique is conducted by (Useche et al., 2019) to collect data for their research that was composed of three core sections: The first part of the questionnaire asked about individual and demographic variables, job-related features and job type and road safety indicators.

Although surveys and self-reports represent a powerful and inexpensive tool for studying various topics in traffic behavior in addition to much of the knowledge in transportation psychology that has been gained by this technique, there is still a dispute regarding the usefulness and validity of such instruments, leading to less than ideal and trustworthy reports on one's own driving behavior and some serious limitations that must be taken into account when using these methods.

3.1.3 Smartphone: The emergence of affordable sensing and computing platforms has a real impact on the appearance of new fields related to driving behavior. One of them is the analysis of driving performance through the use of mobile technology, a field also known as Smartphone Driving Analytics (Carlos et al., 2019). Recently, smartphones have a rich set of on-board sensors such as accelerometers, gyroscopes, GPS, and cameras. These sensors provide valuable information when investigating users' needs and behavioral patterns. Several researchers are currently using mobile phones to collect and gather driving related data. As reported by (Warren et al., 2019) data collection using un-obtrusive technology such as smartphone provides a valuable alternative to study-based data collection. The percentage of 52% (11 out of 21 studies) related to studies that used the mobile phone to acquire data is based only on the use of a cell phone. However, even smartphones are shown to have great potential in data collection. They are largely regarded as dangerous because of its potential to cause distracted driving and crashes.

3.1.4 Traffic surveillances: Road infrastructure development has received widespread attention of many countries in recent years, as well as trying to equip the road with the latest technology. As a result, traffic surveillances instruments have brought new opportunities for researchers in terms of gathering data of driving to investigate different DB's facets. Among researchers who have already taken advantage of this source of data (Zhou et al., 2011) built a framework to define driver behavior patterns by extracting vehicle information from traffic video sequences. Moreover, urban traffic surveillance data at both intersections and road segments were used by (Hongxin et al., 2016) to investigate the driver's involvement in the accident. In addition to a few other researchers addressed the topic of DB using traffic surveillance data, it remains limited use of this source of data.

3.1.5 Datasets: the present time, data is becoming the key for a majority of challenges. Indeed, one-of them is the driving behavior. For many years several researches gathering data related to driving and generating datasets to better understanding the behavior of the driver. (Figure 8) shows some examples of several projects around the world that have collected on-road driving-data according to (Miyajima and Takeda, 2016).

As illustrated previously, the percentage of studies selected for this work that have used pre-collected data sets is 8%. The database of 100-Car Study conducted by the Virginia Tech Transportation Institute was used to modelling of driver Car-Following behavior (Sangster et al., 2013). While (Hamzeie, 2016) investigates how speed limits affect driver speed selection using both data collected on real-time with a Roadway Information Database. (Hallmark et al., 2015) and (Lv et al., 2019) among others several researchers take advantage of the rich Second Strategic Highway Research Program 2 Naturalistic Driving Study (SHRP 2 NDS) datasets to investigate and study assertive approaches of the driving. Furthermore, (Li et al., 2019) used the data set of the electric vehicles to identify the driving patterns.

Project (Period of data collection)	Area	Type of Vehicles	Number of Drivers or Vehicles
100-Car (2003-2004)	United States	Passenger cars	241 drivers 100 cars
SHRP2 (2010-2013)	United States	Passenger cars	3150 drivers
euroFOT (2010-2011)	Europe	Passenger cars, trucks	1,100 drivers
UDRIVE (2014-2016)	Europe	Passenger cars, trucks, powered two-wheelers	290 drivers
400-car (2014)	Australia	Passenger cars	400 drivers
ISAE/TUAT (2005-present)	Japan	Taxis	200 taxis
HQL (2001-2003)	Japan	Instrumented vehicle	97 drivers
NUDrive (1999-2010)	Japan	Instrumented vehicle	800 drivers

Figure 8. Examples of on-road driving-data corpora

3.2 Driver Vehicle Environment's Data

Driving is a driver-vehicle-road environment system and all the three elements affect each other and the whole system. One driver behavior error or vehicle fault or road environment anomaly may lead to another and a chain of reactions within the whole driving system (Mao et al., 2019).

Thus, researches have been interested in these three dimensions of driving behavior for a long time; in order to clearly distinguish the relationship between the different dimensions of

the DB and the DVE model and also to better extract the driving's dimension addressed in each article, we have been based principally on theoretical framework proposed by (Elamrani Abou El Assad et al., 2020).

According to the statistics of the studies selected for this paper (Figure 7), it is clear that researchers are more interested in vehicle-related data collection than in driver-related data or the surrounding environment at rates of 52%, 29% and 19% respectively.

Vehicle related data includes several kinematics such as speed and acceleration/deceleration are the most common measurements in the scientific literature because of its direct impact on the driving behavior and also the opportunity to easily getting them.

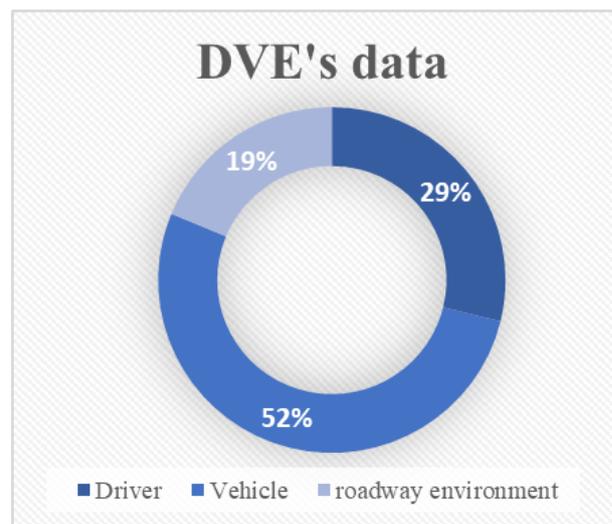


Figure 7. Driver, Vehicle and Environment related data
The driver's profile and state, namely the physiological and psychological conditions provide in its turn relevant and essential information for understanding the driving behavior. In fact, due to the direct contribution of the human being in the driving process, his/her own data have a very strong influence on predicting and detecting of driving events. However, collecting such data usually needs more equipment and sensors.

Finally, surrounding environment data including road geometry, road condition, road type, traffic and the weather condition often requires remote access to the data.

4. CONCLUSIONS

The present survey, in particular the last two sections on data collection and types of data collected, provides a few good insights indicating the need of high-resolution driver's data. On the basis of this, we got the following results:

- Most researchers are interested in the use of in-vehicle sensors to acquire information related to driving behavior. Also, vehicle data have been the primary focus of data collection.
- The percentage of the studies that cover driver's dimension remains very limited, knowing that the driver plays both the role of the controller and the major evaluator of the vehicle quality and the path-following.

This study cannot claim to be complete although we believe that it will be a valuable resource for anyone interested in research on driving behavior in general and data acquisition in particular.

REFERENCES

- Andria G, Attivissimo F, Di Nisio A, Lanzolla AML, Pellegrino A. Design and implementation of automotive data acquisition platform. *Conf Rec - IEEE Instrum Meas Technol Conf 2015;2015-July:272–7*.
<https://doi.org/10.1109/I2MTC.2015.7151278>.
- Boquete L, Rodríguez-Ascariz JM, Barea R, Cantos J, Miguel-Jiménez JM, Ortega S. Data acquisition, analysis and transmission platform for a Pay-As-You-Drive system. *Sensors 2010;10:5395–408*.
<https://doi.org/10.3390/s100605395>.
- Boufous S, Ivers R, Senserrick T, Stevenson M, Norton R, Williamson A. Accuracy of self-report of on-road crashes and traffic offences in a cohort of young drivers: The DRIVE study. *Inj Prev 2010;16:275–7*.
<https://doi.org/10.1136/ip.2009.024877>.
- Bouhoute A, Oucheikh R, Boubouh K, Berrada I. Advanced Driving Behavior Analytics for an Improved Safety Assessment and Driver Fingerprinting. *IEEE Trans Intell Transp Syst 2019;20:2171–84*.
<https://doi.org/10.1109/TITS.2018.2864637>.
Cambridge D. data collection definition n.d.
- Carlos MR, Gonzalez LC, Wahlstrom J, Ramirez G, Martinez F, Runger G. How Smartphone Accelerometers Reveal Aggressive Driving Behavior?--The Key Is the Representation. *IEEE Trans Intell Transp Syst 2019;1–11*.
<https://doi.org/10.1109/tits.2019.2926639>.
- Carvalho E, Ferreira B V., Ferreira J, De Souza C, Carvalho H V., Suhara Y, et al. Exploiting the use of recurrent neural networks for driver behavior profiling. *Proc Int Jt Conf Neural Networks 2017;2017-May:3016–21*.
<https://doi.org/10.1109/IJCNN.2017.7966230>.
- Cen J, Wang Z, Wang C, Liu F. A System Design for Driving Behavior Analysis and Assessment. *Proc - 2016 IEEE Int Conf Internet Things; IEEE Green Comput Commun IEEE Cyber, Phys Soc Comput IEEE Smart Data, IThings-GreenCom-CPSCom-Smart Data 2016 2017:882–7*.
<https://doi.org/10.1109/IThings-GreenCom-CPSCom-SmartData.2016.182>.
- Ding C, Liu M, Wang Y, Yan F, Yan L. Behavior evaluation based on electroencephalograph and personality in a simulated driving experiment. *Front Psychol 2019;10:1–12*.
<https://doi.org/10.3389/fpsyg.2019.01235>.
- Elamrani Abou El Assad Z, Mousannif H. Understanding driving behavior: Measurement, modeling and analysis. vol. 915. Springer International Publishing; 2019.
https://doi.org/10.1007/978-3-030-11928-7_41.
- Elamrani Abou El Assad Z, Mousannif H, Al Moatassime H, Karkouch A. The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. *Eng Appl Artif Intell 2020;87:103312*.
<https://doi.org/10.1016/j.engappai.2019.103312>.
- Hallmark SL, Tyner S, Oneyear N, Carney C, McGehee D. Evaluation of driving behavior on rural 2-lane curves using the SHRP 2 naturalistic driving study data. *J Safety Res 2015;54:17.e1-27*.
<https://doi.org/10.1016/j.jsr.2015.06.017>.
- Hamzeie R. The interrelationships between speed limits , geometry , and driver behavior: a proof-of-concept study utilizing naturalistic driving data 2016.
- Hongxin C, Xin P, Zuo Z, Danya Y, Qiaojun F, Zizhuo W. Driving behavior differences between crash-involved and crash-not-involved drivers using urban traffic surveillance data. *Proc - 2016 IEEE Int Conf Serv Oper Logist Informatics, SOLI 2016 2016:249–54*.
<https://doi.org/10.1109/SOLI.2016.7551696>.
- Julian PT Higgins SG. *Cochrane Handbook for Systematic Reviews of Interventions 2009*.
- Kimberly Madia. *Big data on wheels 2014*.
- Kitchenham B, Charters S. Guidelines for performing Systematic Literature Reviews in SE 2007:1–44.
<https://doi.org/10.1145/1134285.1134500>.
- Kitchenham BA, Mendes E, Travassos GH. Cross versus within-company cost estimation studies: A systematic review. *IEEE Trans Softw Eng 2007;33:316–29*.
<https://doi.org/10.1109/TSE.2007.1001>.
- Li X, Zhang Q, Peng Z, Wang A, Wang W. A data-driven two-level clustering model for driving pattern analysis of electric vehicles and a case study. *J Clean Prod 2019;206:827–37*.
<https://doi.org/10.1016/j.jclepro.2018.09.184>.
- Lv B, Yue R, Zhang Y. The Influence of Different Factors on Right-Turn Distracted Driving Behavior at Intersections Using Naturalistic Driving Study Data. *IEEE Access 2019;7:137241–50*.
<https://doi.org/10.1109/access.2019.2942841>.
- Mao Y, Zhang W, Wang M, Guo D. Driving simulator data based driver behavior ethogram establishment. *1st Int Conf Ind Artif Intell IAI 2019 2019:1–5*.
<https://doi.org/10.1109/ICIAI.2019.8850803>.
- Miyajima C, Takeda K. Driver-Behavior Modeling Using On-Road Driving Data. *IEEE Signal Process Mag 2016:14–21*.
- Sabey BE, Taylor H. *The Known Risks We Run: The Highway. Soc. Risk Assess., 1980*.
https://doi.org/10.1007/978-1-4899-0445-4_3.
- Sangster J, Rakha H, Du J. Application of naturalistic driving data to modeling of driver car-following behavior. *Transp Res Rec 2013:20–33*.
<https://doi.org/10.3141/2390-03>.
- Stanton NA, Salmon PM. Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Saf Sci 2009;47:227–37*.
<https://doi.org/10.1016/j.ssci.2008.03.006>.
- Useche SA, Cendales B, Alonso F, Pastor JC, Montoro L. Validation of the Multidimensional Driving Style Inventory (MDSI) in professional drivers: How does it work in transportation workers? *Transp Res Part F Traffic Psychol Behav 2019;67:155–63*.
<https://doi.org/10.1016/j.trf.2019.10.012>.

Warren J, Lipkowitz J, Sokolov V. Clusters of Driving Behavior from Observational Smartphone Data. *IEEE Intell Transp Syst Mag* 2019;11:171–80. <https://doi.org/10.1109/MITS.2019.2919516>.

Wen J, Li S, Lin Z, Hu Y, Huang C. Systematic literature review of machine learning based software development effort estimation models. *Inf Softw Technol* 2012;54:41–59. <https://doi.org/10.1016/j.infsof.2011.09.002>.

Xie P, Du L, Zhou B, Yu Y, Du H, Cui L, et al. Design and Implementation of Vehicle Data real-time Acquisition System. *J Phys Conf Ser* 2019;1176. <https://doi.org/10.1088/1742-6596/1176/5/052059>.

Yang L, Ma R, Zhang HM, Guan W, Jiang S. Driving behavior recognition using EEG data from a simulated car-following experiment. *Accid Anal Prev* 2018a;116:30–40. <https://doi.org/10.1016/j.aap.2017.11.010>.

Yang L, Ma R, Zhang HM, Guan W, Jiang S. Driving behavior recognition using EEG data from a simulated car-following experiment. *Accid Anal Prev* 2018b;116:30–40. <https://doi.org/10.1016/j.aap.2017.11.010>.

Zhou K, Varadarajan KM, Vincze M, Liu F. Driving behavior inference from traffic surveillance data. *IEEE Conf Intell Transp Syst Proceedings, ITSC* 2011:600–5. <https://doi.org/10.1109/ITSC.2011.6083079>.

Zhu X, Yuan Y, Hu X, Chiu YC, Ma YL. A Bayesian Network model for contextual versus non-contextual driving behavior assessment. *Transp Res Part C Emerg Technol* 2017;81:172–87. <https://doi.org/10.1016/j.trc.2017.05.015>.

APPENDIX

Bellow a comprehensive list of all the 120 selected articles, contains Papers' title, DOI / URL, and results of data extraction

ID	Title	Citation DOI/URL	Source of data				Dimension			
			In-vehicle sensors	Self-report	Other sensors	smartphone	Traffic surv	Dataset	Driver	Vehicle
S1	Driving Behavior Inference from Traffic Surveillance Data	10.1109/ITSC.2011.6083079					✓			✓
S2	Identification of Warning Signs in Car Driving Behavior Before Safety-Critical Events	10.1109/ITSC.2011.6083093	✓		✓				✓	✓
S3	International Large-Scale Vehicle Corpora for Research on Driver Behavior on the Road	10.1109/TITS.2011.2167680	✓		✓				✓	✓
S4	Motion State Acquisition and Error Analysis for Intelligent Vehicle Power Management Strategies	10.1109/ICVES.2011.5983758	✓						✓	✓
S5	Driving Motion Capture Based Driver Behavior Analysis	10.1109/ITSC.2012.6338912			✓					✓
S6	Driver Classification and Driving Style Recognition using Inertial Sensors	10.1109/IVS.2013.6629603	✓							✓
S7	Design and Implementation of Virtual Driving System Fusing Driver's Cognitive and Operating Characteristics	10.1109/ICIEA.2014.6931464			✓					✓
S8	Development of a Data Acquisition and Tracking System for Vehicles	10.1109/EFEA.2014.7059936	✓							✓
S9	Driver Behaviour Profiling using Smartphone Sensory Data in a V2I Environment	10.1109/ICCVE.2014.7297609			✓					✓
S10	Electric Vehicle Data Acquisition System	10.1109/IEVC.2014.7056140	✓		✓					✓
S11	The Study of EV Data Collection and Analysis based on Taiwan i-EV Pilot Project	10.1109/EVS.2013.6914756	✓							✓
S12	A First Look at Vehicle Data Collection via Smartphone Sensors	10.1109/SAS.2015.7133607	✓			✓				✓
S13	An Unsupervised Approach for Inferring Driver Behavior From Naturalistic Driving Data	10.1109/TITS.2015.2449837	✓		✓					✓
S14	Design and implementation of automotive data acquisition platform	10.1109/I2MTC.2015.7151278	✓		✓					✓
S15	Opportunistic Calibration of Smartphone Orientation in a Vehicle	10.1109/WoWMoM.2015.7158210				✓				✓
S16	Theoretical Framework for Physiological Profiling using Sensors and Big Data Analytics	10.1109/CCIntelS.2015.7437952	✓			✓			✓	✓
S17	Driving behavior differences between crash-involved and crash-not-involved drivers using urban traffic surveillance data	10.1109/SOLI.2016.7551696					✓		✓	✓
S18	A System Design for Driving Behavior Analysis and Assessment	10.1109/Things-GreenCom-CPSCCom-SmartData.2016.182	✓	✓						✓
S19	An Intelligent Smartphone Based Approach Using IoT for Ensuring Safe Driving	10.1109/ICECOS.2017.8167137				✓				✓
S20	Driver Assistant for the Detection of Drowsiness and Alcohol Effect	10.1109/SSPS.2017.8071606			✓					✓
S21	Exploiting the Use of Recurrent Neural Networks for Driver Behavior Profiling	10.1109/IJCNN.2017.7966230			✓	✓				✓
S22	SafeDrive: Online Driving Anomaly Detection from Large-Scale Vehicle Data	10.1109/TH.2017.2674661	✓							✓
S23	A Data-Driven Approach for Driving Safety Risk Prediction Using Driver Behavior and Roadway Information Data	10.1109/TITS.2017.2700869						✓	✓	✓
S24	Comparison of Different Driving Style Analysis Approaches based on Trip Segmentation over GPS Information	10.1109/BigData.2017.8258379	✓							✓
S25	Integrating Driving Behavior and Traffic Context through Signal Symbolization for Data Reduction and Risky Lane Change Detection	10.1109/tiv.2018.2843171	✓		✓					✓
S26	Vehicular Data Acquisition and Analytics System for Real-Time Driver Behavior Monitoring and Anomaly Detection	10.1109/ICIINFS.2017.8300417	✓							✓
S27	A Smartphone-based Probe Data Platform for Road Management and Safety in Developing Countries	10.1109/ICDMW.2018.00095				✓				✓
S28	Advanced Driving Behavior Analytics for an Improved Safety Assessment and Driver Fingerprinting	10.1109/TITS.2018.2864637	✓							✓
S29	Analysis of taxi driving behavior and driving risk based on trajectory data	10.1109/IVS.2019.8813887	✓							✓
S30	Driver distraction recognition based on smartphone sensor data	10.1109/SMC.2018.00144	✓			✓				✓
S31	How Smartphone Accelerometers Reveal Aggressive Driving Behavior?—The Key Is the Representation	10.1109/tits.2019.2926639				✓				✓
S32	A graphical modeling method for individual driving behavior and its application in driving safety analysis using GPS data	10.1016/j.trf.2019.03.017	✓							✓
S33	Inference of driver behavior using correlated IoT data from the vehicle telemetry and the driver mobile phone	10.15439/2019F263	✓			✓				✓
S34	A data-driven two-level clustering model for driving pattern analysis of electric vehicles and a case study	10.1016/j.jclepro.2018.09.184	✓							✓
S35	Driving Simulator Data Based Driver Behavior Ethogram Establishment	10.1109/ICIAI.2019.8850803	✓		✓					✓
S36	Social desirability and self-reported driving behaviours: Should we be worried?	10.1016/j.trf.2010.04.004			✓					✓
S37	An investigation of the relationship between the driving behavior questionnaire and objective measures of highway driving behavior	10.1016/j.trf.2012.08.001			✓					✓
S38	Driver behavior in car-to-pedestrian incidents: An application of the Driving Reliability and Error Analysis Method (DREAM)	10.1016/j.aap.2012.05.034	✓		✓					✓
S39	Driving behavior at a roundabout: A hierarchical Bayesian regression analysis	10.1016/j.trd.2013.10.003	✓							✓
S40	Analysis of the Driving Behaviour at Weaving Section Using Multiple Traffic Surveillance Data	10.1016/j.trpro.2014.10.090					✓			✓
S41	Evaluation of driving behavior on rural 2-lane curves using the SHRP 2 naturalistic driving study data	10.1016/j.jsr.2015.06.017					✓	✓	✓	✓
S42	Evaluating the effect of vegetation and clear zone width on driver behavior using a driving simulator	10.1016/j.trf.2016.07.002	✓		✓					✓
S43	The value of self-report measures as indicators of driving behaviors among young drivers	10.1016/j.trf.2016.03.005	✓	✓						✓
S44	Development of an automotive data acquisition platform for analysis of driving behavior	10.1016/j.measurement.2016.07.035	✓		✓					✓
S45	Weather and road geometry impact on longitudinal driving behavior: Exploratory analysis using an empirically supported acceleration modeling framework	10.1016/j.trc.2016.01.017			✓					✓
S46	Detection of driver engagement in secondary tasks from observed	10.1016/j.aap.2017.07.010						✓		✓

	naturalistic driving behavior								
S47	A Bayesian Network model for contextual versus non-contextual driving behavior assessment	10.1016/j.trc.2017.05.015	✓	✓			✓	✓	
S48	A Smartphone Based Technique to Monitor Driving Behavior using DTW and Crowdsensing	10.1016/j.pmcj.2017.06.003			✓			✓	
S49	Evaluation of an in-vehicle monitoring system (IVMS) to reduce risky driving behaviors in commercial drivers: Comparison of in-cab warning lights and supervisory coaching with videos of driving behavior	10.1016/j.jsr.2016.12.008	✓				✓	✓	
S50	Validating the efficacy of GPS tracking vehicle movement for driving behaviour assessment	10.1016/j.tbs.2016.05.001	✓					✓	
S51	Driving Behavior Prediction at Roundabouts Based on Integrated Simulation Platform	10.4271/2018-01-0033	✓		✓			✓	
S52	An online estimation of driving style using data-dependent pointer model	10.1016/j.trc.2017.11.001	✓					✓	✓
S53	Vehicle ownership and other predictors of teenagers risky driving behavior: Evidence from a naturalistic driving study	10.1016/j.aap.2018.06.001	✓	✓				✓	✓
S54	The relation between self-reported driving style and driving behaviour. A simulator study	10.1016/j.trf.2018.04.017	✓	✓					✓
S55	Driver identification in intelligent vehicle systems using machine learning algorithms	10.1049/iet-its.2017.0254					✓		✓
S56	A framework for evaluating aggressive driving behaviors based on in-vehicle driving records	10.1016/j.trf.2017.11.021	✓						✓
S57	Quantifying regional heterogeneity effect on drivers' speeding behavior using SHRP2 naturalistic driving data: A multilevel modeling approach	10.1016/j.trc.2019.06.017					✓	✓	✓
S58	Embedding local driving behaviour in regional emission models to increase the robustness of on-road emission inventories	10.1016/j.trd.2019.05.011	✓	✓				✓	✓
S59	Two-layer pointer model of driving style depending on the driving environment	10.1016/j.trb.2019.08.009	✓						✓
S60	Validation of the Multidimensional Driving Style Inventory (MDSI) in professional drivers: How does it work in transportation workers?	10.1016/j.trf.2019.10.012			✓			✓	✓
S61	The potential of naturalistic driving for in-depth understanding of driver behavior: UDRIVE results and beyond	10.1016/j.ssci.2018.12.029	✓		✓			✓	✓
S62	Analysis of commercial truck drivers' potentially dangerous driving behaviors based on 11-month digital tachograph data and multilevel modeling approach	10.1016/j.aap.2019.105256			✓				✓
S63	Typical-driving-style-oriented Personalized Adaptive Cruise Control design based on human driving data	10.1016/j.trc.2019.01.025	✓		✓			✓	✓
S64	Data Acquisition, Analysis and Transmission Platform for a Pay-As-You-Drive System	10.3390/s100605395	✓		✓				✓
S65	Investigating Differences Between Experienced Adult Drivers and Teen Drivers with Low-Cost Vehicle Data Recorder	10.3141/2321-11			✓				✓
S66	Measuring the Safety Impact of Road Infrastructure Systems on Driver Behavior: Vehicle Instrumentation and Exploratory Analysis	10.1080/15472450.2016.1198699	✓						✓
S67	Studying Driving Behavior on Horizontal Curves using Naturalistic Driving Study Data	10.1177/0361198118784384					✓	✓	✓
S68	Using naturalistic driving data to better understand the driving exposure and patterns of older drivers	10.1080/15389588.2017.1379601	✓		✓			✓	✓
S69	Application of Trajectory Data for Investigating Vehicle Behavior in Mixed Traffic Environment	10.1177/0361198118787364					✓		✓
S70	driver distraction recognition based on smartphone sensor data	10.1109/SMC.2018.00144	✓		✓				✓
S71	Clusters of Driving Behavior From Observational Smartphone Data	10.1109/MITS.2019.2919516			✓			✓	✓
S72	Preliminary establishment and analysis of automated driving benchmark in China	10.1051/mateconf/201925902002	✓		✓				✓
S73	Behavior Evaluation Based on Electroencephalograph and Personality in a Simulated Driving Experiment	10.3389/psyg.2019.01235			✓	✓			✓
S74	Testing the Effectiveness of Objective and Subjective Predictors of Driving Behavior under Fatigue	10.1177/0361198119843099	✓	✓	✓				✓
S75	DBNet: A Large-Scale Dataset for Driving Behavior Learning	http://www.dbehavior.net/data/egpa_per_release.pdf	✓		✓				✓
S76	Traffic Data Collection and its Standardization	10.1007/978-1-4419-6070-2					✓		✓
S77	Fatigue Driving Detection System Design Based on Driving Behavior	10.1109/ICVIP.2010.101	✓		✓			✓	✓
S78	Effects of Individual Differences on Driving Behavior and Traffic Flow Characteristics	10.3141/2248-01	✓	✓	✓	✓		✓	✓
S79	A Driving Behavior Retrieval Application for Vehicle Surveillance System	10.5815/ijmecs.2011.02.07	✓					✓	✓
S80	Mixed effects model on driver steering behavior under emergency	10.1109/CCDC.2011.5968980	✓		✓			✓	✓
S81	Comparison on Driving Behavior between Manned and Unmanned Ground Vehicles	10.1109/ICICIP.2011.6008299	✓		✓			✓	✓
S82	Using Mobile Phone Sensors to Detect Driving Behavior	10.1145/2442882.2442941					✓		✓
S83	Longitudinal Driving Behavior with Integrated Crash-Warning System	10.3141/2365-03	✓		✓			✓	✓
S84	Application of Naturalistic Driving Data to Modeling of Driver Car-Following Behavior	10.3141/2390-03	✓		✓			✓	✓
S85	Design and Implementation of Driving Behavior Analysis System	10.4304/jnw.9.8.2114-2120	✓	✓	✓				✓
S86	A Study of Individual Characteristics of Driving Behavior Based on Hidden Markov Model	https://pdfs.semanticscholar.org/0b6a/09d5dcade342ca3795f2bfd5d8f17469933.pdf			✓			✓	✓
S87	Vehicle-Based Driver Interlocks: After 50 Years of Research, Can We Now Design an On-Board, Real-Time System to Detect Abnormal Driving Behavior?	10.4271/2014-01-0275	✓		✓				✓
S88	Evaluation method for aggressiveness of driving behavior using drive recorders	10.1541/ieejia.4.59	✓		✓				✓
S89	Localization and Driving Behavior Classification with Smartphone Sensors in Direct Absence of Global Navigation Satellite Systems	10.3141/2489-08					✓		✓
S90	Driver-Behavior Modeling Using On-Road Driving Data	10.1109/MSP.2016.2602377	✓		✓			✓	✓
S91	The interrelationships between speed limits, geometry, and driver behavior: a proof-of-concept study utilizing naturalistic driving data	10.31274/etd-180810-4594	✓		✓			✓	✓
S92	A Novel Model-Based Driving Behavior Recognition System Using Motion Sensors	10.3390/s16101746	✓		✓			✓	✓
S93	Is There Any Relationship between Mental Health and Driving Behavior of Taxi Drivers in Kerman?	10.5539/ghs.v9n2p294			✓				✓
S94	Smart Driving: Influence of Context and Behavioral Data on Driving Style	10.1007/978-3-319-46301-8_12			✓			✓	✓
S95	Computational Collective Intelligence	10.1007/978-3-319-67077-5					✓		✓
S96	Deep Learning based Traffic Direction Sign Detection and Determining Driving Style	10.1109/UBMK.2017.8093453					✓		✓
S97	Using OBD-II Data to Explore Driving Behavior Model	10.1109/ICASI.2017.7988297	✓						✓
S98	Syntactic Pattern Recognition of Car Driving Behavior Detection	10.1145/302227.3022303					✓		✓

S99	Research on the Classification and Identification of Driver's Driving Style	10.1109/ISCID.2017.31	✓						✓
S100	Toward Reasoning of Driving Behavior	10.1109/ITSC.2018.8569250	✓	✓					✓
S101	Investigation of driver behavior during crash and near-crash events using naturalistic driving data	https://lib.dr.iastate.edu/etd/16556	✓	✓			✓	✓	✓
S102	In-Vehicle Affect Detection System: Identification of Emotional Arousal by Monitoring the Driver and Driving Style	10.1145/3239092.3267417	✓	✓	✓			✓	✓
S103	Big Data Research on Driving Behavior Model and Auto Insurance Pricing Factors Based on UBI	10.1007/978-981-10-7521-6_49	✓	✓	✓	✓	✓	✓	✓
S104	Assessing the impacts of driving environment on driving behavior patterns	10.1007/s11116-018-9965-5	✓	✓	✓		✓	✓	✓
S105	Investigation of drivers speed selection behavior: A naturalistic driving approach	https://lib.dr.iastate.edu/etd/16589	✓	✓			✓	✓	✓
S106	Driving behavior recognition using EEG data from a simulated car-following experiment	10.1016/j.aap.2017.11.010	✓					✓	✓
S107	Unobtrusive Driver Drowsiness Prediction using Driving Behavior from Vehicular Sensors	10.1109/ICPR.2018.8545427	✓					✓	✓
S108	Driving Style Recognition using Deep Learning Approach	http://universalreview.org/gallery/22-feb2019.pdf				✓			✓
S109	Identifying Important Risk Factors Associated with Vehicle Injuries Using Driving Behavior Data and Predictive Analytics	10.1109/ICHI.2019.8904860						✓	✓
S110	Analysis of cut-in behavior based on naturalistic driving data	10.1016/j.aap.2019.01.006	✓	✓				✓	✓
S111	Research on safe driving behavior of transportation vehicles based on vehicle network data mining	10.1002/ett.3772	✓						✓
S112	Modeling Driving Behavior at Single-Lane Roundabouts	10.24355/dbbs.084-201903071236-0	✓					✓	✓
S113	Location-Based Analysis of Car-Following Behavior During Braking Using Naturalistic Driving Data	10.1139/cjce-2019-0314						✓	✓
S114	The Influence of Different Factors on Right-Turn Distracted Driving Behavior at Intersections Using Naturalistic Driving Study Data	10.1109/access.2019.2942841						✓	✓
S115		10.3389/psyg.2019.01235				✓			✓
S116	Are you a Good Driver? A Data-driven Approach to Estimate Driving Style	10.1145/3307363.3307375	✓	✓	✓				✓
S117	Rapid Driving Style Recognition in Car-Following Using Machine Learning and Vehicle Trajectory Data	10.1155/2019/9085238					✓	✓	✓
S118	Driving Style Recognition Based on Electroencephalography Data From a Simulated Driving Experiment	10.3389/psyg.2019.01254	✓	✓				✓	✓
S119	Driving style estimation by fusing multiple driving behaviors: a case study of freeway in China	10.1007/s10586-018-1739-5	✓	✓					✓
S120	Fine-grained Abnormal Driving Behaviors Detection and Identification with Smartphones	10.1109/TMC.2016.2618873					✓		✓