Enhancing the Architecture of Context-Aware Driver Assistance Systems by Incorporating Insights from Naturalistic Driving Data

by

Mostafa Mohammed Hasan Tawfeek Mohammed

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Transportation Engineering

Department of Civil and Environmental Engineering
University of Alberta

ABSTRACT

Driving assistance systems (DASs) have received a great deal of attention in the past decades as an active and effective collision countermeasure. DASs potential benefits will be attained by enhancing the systems' awareness regarding the dynamic driving context including the change in the driver behavior, vehicle status, and surrounding environment. Therefore, context-aware systems were proposed to improve the adaptability of the DASs to that dynamic driving context. This dissertation proposes a new aspect to the architecture of the context-aware DASs by introducing a context identification layer to the reasoning subsystem. The main goal of this layer is to accommodate changes in driver behavior due to the surrounding environment and to customize the system to the needs of each individual driver. The proposed layer is designed based on key insights that were drawn from naturalistic driving studies based on actual driving behavior. Data from 64 drivers in a Naturalistic Driving Study (NDS) was selected as a measurement of driver behavior because it provides network-wide data collection without restrictions or instructions to the driver on the routes that should be used or the roadway elements (e.g., curves or two-way two-lane roads), making the data collected closest to reality. The data was processed, and events of interest were extracted, mapped using ArcGIS, and analyzed to differentiate between the intersection- and segment-related events. The intersection-related events were identified according to the intersection influence area, which was estimated based on the stopping sight distance and the speed limit. Several behavioral measures were extracted for each event of interest, including following distance, relative speed, headway between the host vehicle and the leading vehicle, acceleration, time-to-collision, and jerk of the host vehicle.

Based on the insights observed, the proposed context identification layer contains two algorithms that work in sequence: the infrastructure detection algorithm and the driver classification

algorithm, respectively. The infrastructure detection algorithm aims at identifying intersection-related driving where the driver adjusts his or her behavior due to the presence of an intersection ahead. This algorithm was developed by training a deep neural network for each driver to accommodate each driver needs. The output of this was then used to initiate the processing of the driver classification algorithm. The driver classification algorithm categorizes drivers into cautious, normal, and aggressive drivers using the first algorithm (i.e., near intersections or on segments) and combining behavioral measures using unsupervised machine learning techniques including principal component analysis and *K*-means. The provided key insights into driver behavior included, but was not limited to: (i) the quantification of driving behavior in proximity to intersections and on segments in the form of probability density functions, (ii) there were significant differences between both behaviors in terms of the measures used, (iii) the proposed definition of the intersection influence area was plausible in terms of distinguishing between the segment- and intersection-related driving behavior, (iv) acceleration was the highest among the behavioral measures influenced by proximity to an intersection, and (v) the driver classification, which ignores the driver's relative location to intersections, were more likely to misclassify drivers as aggressive when they were in high intersection density areas such as downtown cores.

PREFACE

The chapters of this dissertation are published or presented in peer-reviewed journals or conferences. Moreover, there are two manuscripts under review. These articles are as follows:

1. **Tawfeek, M. H.** and K. El-Basyouny (2018) A Perceptual Forward Collision Warning Model using Naturalistic Driving Data. *Canadian Journal of Civil Engineering.*
2. **Tawfeek, M. H.** and K. El-Basyouny (2018) Using Naturalistic Driving Data to Quantify Driver Following Behavior During Braking. *Proceedings, Annual Conference - Canadian Society for Civil Engineering*.
3. **Tawfeek, M. H.**, and K. El-Basyouny. A Context-aware Driving Assistance System using Naturalistic Driving Data. (*under review*)
4. **Tawfeek, M. H.**, and K. El-Basyouny. Location-based Driver Classification using Naturalistic Driving Data. (*under review*)

To my dear parents, Mohammed and Magda

my beloved wife, Sarah

my sweet daughter, Leen

ACKNOWLEDGMENTS

"*…And my success is not but through Allah. Upon him I have relied, and to Him I return"* The Holy Quran 11:88.

All praise is due to Allah the All-Knowing and the Most Merciful for helping me to complete this work successfully.

I also express my gratitude to my supervisor, Prof. Karim El-Basyouny for his time, and advice. I would like to offer my thanks to my examination committee, Prof. Amy M. Kim, Prof. Tae J. Kwon, Prof. Yasser Mohamed, and Prof. Mohamed Ahmed for their time and comments.

I would like also to thank Prof. Mohamed El Esawey for his support and advice during my program. I wish to thank my dear friends, Eng. Mahmoud Yassin and Mr. Abdelrahman Hasan.

I would also like to thank all my friends and fellow students in the transportation group.

Also, I would like to offer my thanks to the Faculty of Graduate Studies and Research (FGSR) staff for their support during my internship in the FGSR. I would like to thank Prof. John Nychka, the Associate Dean of the FGSR, for his trust and support during my Graduate Teaching and Learning research project. Also, I am thankful to Medha Samarasinghe, my direct supervisor during my internship in the FGSR, for her support and encouragement.

Lastly but most importantly and before any person, I would like to present my sincere thanks and gratefulness to my father, Eng. Mohammed Hasan, and my mother, Eng. Magda Mansour, for their unconditional and unlimited support during my journey of completing this work. With all my love, I would also like to thank my wife, Dr. Sarah Gemeah, for her continuous and infinite support and encouragement through the tough times that we overcame together during the compilation of this work. Also, I extend my thanks to my dear sister, Dr. Marwa, and my dear brother, Eng, Moaz, for their support.

TABLE OF CONTENTS

Abstract ii

Preface iv

Dedication v

Acknowledgments vi

Table of Contents vii

List of Figures x

List of Tables xii

List of Abbreviations xiii

1. Introduction 1

1.1 Background 1

1.1.1 Context-aware Driver Assistance System 1

1.1.2 DAS Supporting Mediums 2

1.1.3 Driver Behavior 5

1.2 Problem Statement 6

1.3 Research Gaps 7

1.4 Dissertation Objectives 9

1.5 Research Methodology 10

1.6 Thesis Structure 12

2. Literature Review ... 14

2.1 Context-aware DAS ... 14

2.2 Car-Following Behavior During Braking ... 20

2.3 The Detection of Drivers' Behavior Variation ... 26

2.3.1 Drivers' Variation in Simulation Environments ... 26

2.3.2 Drivers' Variation in Test Tracks and On-Road Experiments ... 29

2.3.3 Drivers' Variation using Smartphones ... 31

2.3.4 Drivers' Variation in Naturalistic Driving Environments ... 33

2.4 Summary ... 37

3. Events Extraction and Behavioral Measures Quantification ... 42

3.1 Driver Behavior Measurement Methods ... 42

3.2 NDS and Driver Behavior Studies ... 50

3.3 NDS Data Description ... 56

3.4 Identifying of Car-Following during Braking Events ... 57

3.5 Intersection Influence Area Definition and Events Mapping ... 62

3.6 Behavioral Measures Quantification ... 68

3.7 Summary and Discussion ... 76

4. An Overview on the Context-aware System and The Infrastructure Detection Algorithm ... 78

4.1 Context-Aware DAS Architecture ... 78

4.2 The Proposed Context-Aware Architecture ... 80

4.3 Neural Network Model for the Infrastructure Identification Layer 84

4.3.1 Artificial Neural Networks and Driver Behavior... 84

4.3.2 Individualized Artificial Neural Networks for Driver Behavior Classification...... 87

4.4 Selecting the Optimum Neural Network Size 91

4.5 The Influence of Input Variables on the Neural Network Models................................ 98

4.6 Summary and Discussion 105

5. The Location-Based Driver Behavior Classification Algorithm 108

5.1 Driver Behavior Classes.. 108

5.2 Drivers Classification Procedure.. 111

5.3 Location Classification.. 117

5.4 Behavioral Classification 120

5.5 Summary and Discussion 128

6. Summary 131

6.1 Conclusions 131

6.2 Contributions.. 134

6.3 Future Research and Limitations... 136

References 139

LIST OF FIGURES

Figure 1. Automation levels according to SAE, J3016 (SAE International, 2014)........................ 4

Figure 2. A summary of the research methodology. 12

Figure 3. The relation between different behavior measurement methods according to network-wide measurement and closeness to reality. 49

Figure 4. The main parameters of a car-following situation. 60

Figure 5. Events of interest and behavioral measures extraction framework.............................. 61

Figure 6. Sketch of intersection influence area. 63

Figure 7. Example of the intersection events (red dots) and segment events (blue dots). 64

Figure 8. The distribution of intersection (11403 events) (red dots) and segment (14962 events) (blue dots) car following events while braking in the low-speed group..................................... 66

Figure 9. The distribution of intersection (1995 events) (red dots) and segment (12614 events) (blue dots) car following events while braking in the medium speed group. 66

Figure 10. The distribution of segment car following events (3409 events) (blue dots) while braking in the high-speed group. 67

Figure 11. Probability Density Functions for the acceleration at different locations (speed category). 74

Figure 12. Probability Density Functions for the following distance at different locations (speed category). 74

Figure 13. Probability Density Functions for the TTC at different locations (speed category)... 75

Figure 14. Probability Density Functions for the relative speed at different locations (speed category). 75

Figure 15. Probability Density Functions for the jerk at different locations (speed category). ... 76

Figure 16. The proposed context-aware DAS architecture. 82

Figure 17. A summary of the context identification layer decisions. 83

Figure 18. An example of the 3-hidden layers neural network model structure. 88

Figure 19. The frequency of the drivers with respect to the number of the events. 90

Figure 20. The average classification accuracy (Acc.) of the 63 drivers for each neural network size. 93

Figure 21. An example of the classification results for different neural network sizes. 95

Figure 22. Summary of the total accuracies for all the drivers. 97

Figure 23. Summary of the training (on the left) and test (on the right) accuracies for all the drivers. 97

Figure 24. The aggregation of the extracted events. 112

Figure 25. Events aggregation and driver classification framework. 113

Figure 26. The *K*-means clustering results for segment- and intersection-related events. 119

Figure 27. Behavior classification results with the drivers' IDs for case a. 126

Figure 28. Behavior classification results with drivers' IDs for case b. 126

Figure 29. Behavior classification results with drivers' IDs for case c. 127

Figure 30. Contour map for the TTC values for case a. 127

Figure 31. Contour map for the TTC values for case b. 128

Figure 32. Contour map for the headway values for case c. 128

LIST OF TABLES

Table 1 A comparison between different driver behavior measurement methods 47

Table 2 Events breakdown by event type and speed group 65

Table 3 The parameters of the fitted distributions 73

Table 4 Total accuracies for different input variables combinations.. 99

Table 5 Driver behavior classes summary 109

Table 6 The results of the PCA of the data matrix $X_{all\text{-}split}$ 117

Table 7 The loadings for the first two Principal Components of the data matrix $X_{all\text{-}split}$.......... 118

Table 8 The results of the PCA of the data matrices X_{all}, X_{seg}, and X_{int}.................................... 123

Table 9 The loadings for the first two Principal Components of X_{all}, X_{seg}, and X_{int}.................. 124

Table 10 Clustering results of drivers classification... 125

LIST OF ABBREVIATIONS

AASHTO	American Association of State Highway and Transportation Officials
AIC	Akaike Information Criterion
AV	Autonomous Vehicles
BIC	Bayesian Information Criterion
CAN	Controller Area Network
CV	Connected Vehicles
DAS	Driver Assistance System
DSRC	Dedicated Short-Range Communication
DVI	Driver Vehicle Interface
ETTC	Enhanced Time to Collision
FCW	Forward Collision Warning
FHWA	Federal Highway Administration
HSM	Highway Safety Manual
I2V	Infrastructure-to-Vehicle
ISD	Integrated Safety Devices
KS-TEST	Kolmogorov-Smirnov test
LCW	Lane Change Warning
MANET	Mobile ad hoc Network
NDS	Naturalistic Driving Study
NGSIM	Next Generation Simulation
NHTSA	National Highway Traffic Safety Administration

OBE	On-Board Equipment
PC	Principal Component
PCA	Principal Component Analysis
RSU	Road Side Unit
SPaT	Signal Phase and Timing
SPMD	Safety Pilot Model Deployment
SSD	Stopping Sight Distance
TTC	Time to Collision
TET	Time Exposed Time-to-Collision
TIT	Time Integrated Time-to-Collison
V2I	Vehicle-to-Infrastructure Communication
V2V	Vehicle-to-Vehicle Communication
V2X	Vehicle-to-other modes Communication
VANET	Vehicular Ad-hoc Networks
WHO	World Health Organization

1. INTRODUCTION

Every year, approximately 1.24 million deaths and 20 to 50 million non-fatal injuries are recorded worldwide due to road collisions (WHO, 2013). In Canada, more than 1.3 million collisions were reported from 2004 to 2014, and more than 1.7 million injuries and 23,279 fatalities were recorded as a result of road collisions (Transport Canada, 2018). Although there are a number of contributing factors, several studies consider driver (i.e., human) error as the leading cause of about 90% of collisions (Sayed et al., 1995; Treat et al., 1979). Consequently, the automotive industry was fast to react by developing driver assistance systems that integrates communication and sensing technologies to provide active collision countermeasures.

1.1 BACKGROUND

This section will provide a quick overview of some fundamental concepts used in this dissertation.

1.1.1 Context-aware Driver Assistance System

Driver assistance system (DAS) is a system that helps the driver to actively improve his or her safety. One example of a DAS is the Forward Collision Warning (FCW) system, which assists the driver to navigate away from a critical rear-end event by relaying a warning message about an impending rear-end collision. Another example is the Lane Change Warning (LCW) system, which assists the driver by checking the blind spots during a lane change and displaying a warning when a vehicle is detected. Recently, DAS was introduced as context-aware systems that could monitor and detect the change in any of the driving context components (e.g., driver, vehicle, and environment) (Vahdat-Nejad et al., 2016).

A context-aware system was defined by Schilit et al. (1994) as a system that considers the changes in location of use, surrounding people, and accessible devices, in addition to, the change of these elements over time. These systems were also defined as applications that adapt to the surrounding environment based on the user or equipment location and the capabilities of the networking infrastructure (Harter et al., 2002). These definitions of context-aware systems were adopted for DAS. For instance, a context-aware DAS was defined as the system that constantly perceives, interprets, and reacts to the changing driving situations and the variation in the road conditions (Vahdat-Nejad et al., 2016). Also, Boyraz and Hansen (2009) described this system in two main modules, namely, instant driving context identification and deviation detection. These two modules identify three main patterns, namely, context recognition (i.e., identified in terms of maneuver recognition), abnormal behavior detection, and driver distraction level prediction. Based on this information, context-aware DAS could be operated in various mediums and environments, such as Connected Vehicles (CV), Vehicular Ad-hoc Networks (VANET), and Automated Vehicles (AV).

1.1.2 DAS Supporting Mediums

For the DAS to fulfill its function, communication and sensing technologies are needed to collect relevant data. These communication and sensing technologies are integrated into various Intelligent Transportation Systems (ITS) architecture, such as CV, VANET, and AV.

The CV technology allows information sharing among vehicles (V2V), between vehicles-and-infrastructure (V2I) (e.g., traffic control devices), or among vehicles and other modes (V2X) like pedestrians and bicycles, using various communication technologies such as dedicated short-range communication (DSRC). These types of communication feed information to the DAS to initiate the execution of a particular assistance action (i.e., display a warning or perform a specific maneuver).

The basic components of the V2V system are On-Board Equipment (OBE) and Driver Vehicle Interface (DVI). The OBE broadcasts the host vehicle's information (e.g., vehicle position, speed, etc.) on a DSRC channel in the 5.9GHz band, specified exclusively for automotive usage (Kenney, 2011). In addition, the OBE receives similar information from other vehicles within communication range equipped with OBEs, and can process vehicle's information and provide advice or warnings to the driver of the host vehicle through the DVI (Harding et al., 2014). The DVI is often a device that is built-in and fully integrated into the vehicle by the manufacturer, such as screen display or LEDs and blinkers installed in the driver's field of view. The warning message could also be relayed to the driver in various forms such as audio, visual, haptic or a combination of these displays (Harding et al., 2014).

The V2I system primarily consists of Road Side Unit (RSU), OBE, and DVI. At an intersection, for example, the RSU collects the Signal Phase and Timing (SPaT) information from the traffic signal controller and encapsulates as well as encodes it into a standard SPaT message (SAE, 2013). The SPaT message, designed explicitly for infrastructure-to-vehicle (I2V) communication, is then broadcasted on the DSRC channel. The OBE receives the broadcast of the SPaT message and can consequently transmit the host's vehicle information as well while in the communication range of the RSU. The DVI has the same role as mentioned in the V2V system.

Similar to the CV technology, VANET depends on wireless communication among vehicles and between vehicles and the infrastructure. The VANET is considered an ITS communication medium that has a variety of applications in the form of context-aware DAS (Vahdat-Nejad et al., 2016).

On the other hand, AV technology depends mainly on sensing technology in the vehicle that allows different automation levels, and has a wide range of driving automation levels, starting from no automation (level 0) to full automation (level 5) (SAE International, 2014). Figure 1 shows the definitions for each automated level. It is worth mentioning that in this dissertation the Autonomous Vehicles (i.e., self-driving vehicles) is meant to be at the full automation level (level 5) of the AV. The adapted DAS definition in this dissertation could contribute to several levels of automation starting from level 0 and up to level 3, since the context-aware DAS provides the driver with information regarding the driving environment and could take over some driving tasks. However, the main focus of this dissertation is on car-following situations only.

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/ Deceleration	*Monitoring* of Driving Environment	Fallback Performance of *Dynamic Driving Task*	System Capability *(Driving Modes)*
***Human driver* monitors the driving environment**						
0	No Automation	the full-time performance by the *human driver* of all aspects of the *dynamic driving task*, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the *driving mode*-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the *driving mode*-specific execution by one or more driver assistance systems of both steering and acceleration/ deceleration using information about the driving environment and with the expectation that the *human driver* perform all remaining aspects of the *dynamic driving task*	**System**	Human driver	Human driver	Some driving modes
***Automated driving system* ("system") monitors the driving environment**						
3	Conditional Automation	the *driving mode*-specific performance by an *automated driving system* of all aspects of the dynamic driving task with the expectation that the *human driver* will respond appropriately to a *request to intervene*	System	**System**	Human driver	Some driving modes
4	High Automation	the *driving mode*-specific performance by an automated driving system of all aspects of the *dynamic driving task*, even if a *human driver* does not respond appropriately to a *request to intervene*	System	System	**System**	Some driving modes
5	Full Automation	the full-time performance by an *automated driving system* of all aspects of the *dynamic driving task* under all roadway and environmental conditions that can be managed by a *human driver*	System	System	System	**All driving modes**

Figure 1. Automation levels according to SAE, J3016 (SAE International, 2014).

1.1.3 Driver Behavior

The main goal of any DAS is to assist the driver during the driving process. Thus, it is crucial to understand drivers' behavior. Driver behavior is the actual act the driver does such as driving at a certain speed (Evans, 1991). To study normal driver behavior or the effect of DAS on this, different behavioral measures and indicators, obtained from several data collection methods such as driving simulators, controlled test tracks, naturalistic driving studies, etc., have been explored in the literature.

For instance, Mohammed et al. (2016) compared the behavior with and without a DAS warning message in an on-road test using various behavioral measures such as speed profile, acceleration profile, and maximum speed. Vehicle position data and the behavioral measures were collected using a smartphone's navigation system and accelerometer. In addition, Farah et al. (2012) assessed drivers' behavior in terms of driving speed, lane-changing frequency, following gaps, and acceleration noise. CAN-bus, cameras, and LiDAR sensors were used to collect behavioral data in an on-road test. Zhao et al. (2016) used aggregate driver behavioral measures such as maximum acceleration, maximum deceleration, and maximum speed to evaluate an eco-driving system on a corridor segment with multiple signalized intersections. In work zones, Debnath et al. (2017) studied the influence of portable traffic control devices on driver behavior. The behavioral measures included compliance to stop conditions and speed change before and after the devices for all vehicles. For the first vehicle approaching the portable traffic control device, approach speed and deceleration rate, and stopping distance with respect to the control device were used. The data was collected and extracted from video images and pneumatic tubes. In a different study, the effect of roadside vegetation densities and clear zone widths were studied in terms of drivers' operating speed and the ability to maintain their lateral position in a driving simulator (Fitzpatrick et al., 2016).

Generally, driver behavior is the outcome of maneuvers (i.e., braking, steering, etc.). The most common of which are speed, longitudinal and lateral acceleration, lane changing frequency, following gap, etc. It is preferable to use some means over others to investigate certain aspects of driver behavior. For example, naturalistic driving data is preferred to study normal driver behavior over driving simulators or test tracks (Evans, 1991).

1.2 PROBLEM STATEMENT

DASs are expected to have a significant role in crash avoidance, while offering a wide range of mobility and environmental improvements. The expected benefits from DASs will be mainly realized by enhancing the driver's awareness regarding their dynamic driving context, which is a safety-critical task that might not necessarily lead to direct safety improvements.

For example, a DAS can enhance driver awareness by alerting them of safety-critical events through in-vehicle messaging. Consequently, as a warning system, a DAS could generate false or missed alarms. Missed alarms is a system's failure to issue a warning when such a warning is warranted, while false alarms occur when the system relays an unwarranted warning or, in other words, when the situation is not critical from driver's perspective (Bliss and Acton, 2003; Hirst and Graham, 1997). Both missed and false alarms will have a direct effect on drivers' compliance and acceptance as well as the reliability of the warning system (Naujoks et al., 2016a). Moreover, drivers may decide to turn off or ignore the warning if the system consistently relays unreliable information (Tijerina and Garrott, 1997). On the other hand, reliable warnings will accelerate the driver's decision-making process, which can lead to positive impacts (Ruscio et al., 2015).

One important source of false alarms in any warning system is the false detection of the drivers' evasive or safety critical maneuvers. For instance, a frequent number of false alarms was produced, when using acceleration as a measure for critical maneuvers detection (McLaughlin et al., 2008). The evaluation of warning systems showed that warnings too early, reduce users' acceptance and decrease the impact of the warning on driver behavior (Staubach et al., 2014; van Driel et al., 2007). Conversely, in the case of late warnings, drivers could start a maneuver before the warning is relayed, and hence, drivers will develop a mistrust due to the conflict between their expectations and the warning system performance (Abe and Richardson, 2005). Generally, developing an efficient warning system should be based on the knowledge of unassisted driving behavior to avoid the above-mentioned issues (Berndt et al., 2007). Therefore, clear insights into driver's behavior are crucial for improving the detection of critical safety maneuvers, which should ultimately lead to the development of more reliable and driver-accepted warning systems, thereby, maximizing the full potential of context-aware DASs.

1.3 Research Gaps

Considering the importance of understanding driver behavior, as discussed in the previous section, the development of a context-aware DAS based on this has received a great deal of attention in the past decade. However, there are no clear insights into the following aspects of driver behavior and, specifically, when developing these context-aware DAS:

- Car-following is the most frequent driving situation that occurs on a daily basis, given that the continuous increase in light vehicle ownership. Hence, acknowledging this aspect of driver behavior is crucial for the development of context-aware DAS. Despite the extensive research conducted in this area of behavior (Saifuzzaman and Zheng, 2014), a limited number of studies focus on car-following while braking; however, braking is the most frequent response to a car-following critical safety situation (Adams, 1994; Lee et al., 2007). Moreover, there is a lack of previous research on quantifying and comparing car-following behavior while braking at different locations of the road network (e.g., intersections).

- Despite the wide range of proposed context-aware DAS in the literature, these systems either focus on driving data from the host vehicle only (e.g., based on smartphone data), or driving data from the host and surrounding vehicles without considering the environmental component of the driving context. Specifically, these systems did not take into account the effect of the surrounding infrastructure (e.g., intersection) on the driver behavior. Hence, the adaptability of these context-aware DASs is questionable.
- Due to the behavioral variations among drivers, the context-aware DAS warning messages could be perceived by drivers as early or late warnings. For instance, an FCW message could provide an early warning for a driver who tends to follow the leading vehicle too closely. Therefore, the literature highlighted the importance of considering the driver's individual preferences and needs when developing DAS (Bengler et al., 2014). Although several studies investigated the classification of drivers into cautious, normal, and aggressive based on various behavioral measures to enhance the development of context-aware DAS, these studies did not explore the impact the drivers' location (i.e., near intersections) has on the classification.
- There is no link established in the literature between the aggregate driver behavior and their behavior in specific locations. For instance, some studies analyzed specific behavior (e.g., aggressive driving) at particular locations on the roadway, yet these studies did not provide a link between behavior at these locations and driver behavior in general. In other words,

if drivers tend to be more cautious or more aggressive at intersections than normal, it was not clear from the literature whether this was due to the characteristics of the intersections. On the other hand, other studies analyzed the driver classification on an aggregate basis without linking the driver behavior at specific locations (e.g., intersections) to the aggregate behavior, especially for studies that took place in high intersection density areas.

1.4 DISSERTATION OBJECTIVES

The primary goal of this dissertation is to propose a context-aware DAS architecture that acknowledges the variation in driver behavior due to a change in surrounding environment and drivers' individual needs by investigating car-following behavior during braking. To reach this primary goal, and based on the research gaps discussed in the previous section, the following points summarize the tasks of the dissertation:

- Quantify behavioral measures for car-following during braking events at different locations on the road network, using measurement methods that facilitate network-wide behavior data collection.
- Investigate the differences between behavioral measures at a specific location on the network (i.e., near intersections) when compared to other locations (i.e., midblock segments).
- Integrate these differences into the design of context-aware DASs by proposing an additional layer to the system (i.e., context identification layer), which will detect a change in the driving context.
- Develop an individualized model (i.e., for each driver) to detect a change in behavior based on the proximity to an intersection as the first part of the proposed context identification layer (i.e., infrastructure detection algorithm).

- Study the influence the behavioral measures have on the classification accuracy of the infrastructure detection algorithm.
- Develop a driver classification algorithm, as the second part of the proposed context identification layer, that classifies drivers using several behavioral measures extracted from normal driving behavior.
- Examine whether the driver classification of cautious, normal, and aggressive changes due to the location on the road network.

1.5 Research Methodology

In order to achieve the objectives of this dissertation, the workflow summarized in Figure 2 was followed. The methodology started with reviewing the current literature to identify gaps in the current research. Because of the importance of developing a context-aware DAS based on driver behavior, car-following behavior was reviewed with more emphasis on research related to the actions in question during braking. To acknowledge the integration of varying driver behavior into the design of the context-aware DAS, the current state-of-the-art research related to driver behavior variation was also reviewed. In the next stage of the methodology, the typical driver behavior measurement methods including driving simulators, test tracks and on-road experiments, and naturalistic driving studies were reviewed. The naturalistic driving studies was selected as the most suitable method to use. Then the selected method data was obtained and analyzed to extract car-following events during braking with the corresponding behavioral measures. The extraction was followed by pairing the events of interest with the road network maps to differentiate driver behaviors in different locations on the road network. Finally, the events of interest were analyzed, and insights were observed.

Based on these insights, an enhancement was proposed to the architecture of the context-aware DAS in the form of an additional layer in the reasoning subsystem. This proposed layer consists of infrastructure detection algorithm and driver classification algorithm. The infrastructure detection algorithm was developed to detect when the driver will be influenced by an intersection ahead based on a group of behavioral measures using deep learning (i.e., deep neural networks). Moreover, an extensive analysis was performed to highlight the measures that could influence the detection accuracy.

Lastly, a driver classification algorithm was proposed for the location-based driver classification after the infrastructure detection algorithm. The algorithm classified the drivers into cautious, normal, and aggressive based on six of the behavioral measures using machine learning. The machine learning technique combined the principal component analysis with the *K*-means clustering. The principal component analysis was used to extract the maximum information from the measures and, the *K*-means clustering then used the principal components to classify the drivers into three classes. The classification was carried out based on the location of the driver on the road network, and output of the classification was compared between the different locations.

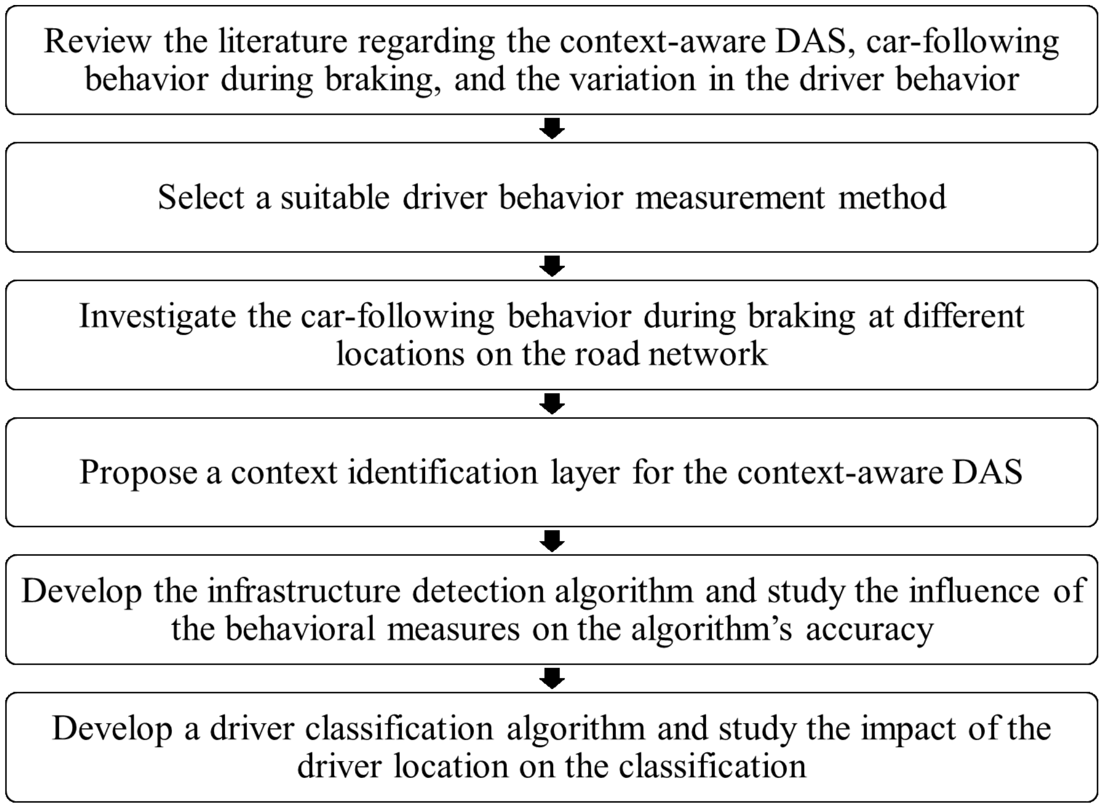


Figure 2. A summary of the research methodology.

1.6 Thesis Structure

Chapter One, presents a background of the dissertation and identifies the problem statement, research gaps, objectives, and the contribution of the dissertation. **Chapter Two** reviews the literature that is related to the context-aware DAS, car-following behavior, and the variation in drivers' behavior. The car-following behavior is discussed in three phases, namely, pre-, during, and post-braking with more attention to the during braking phase. The variation in drivers' behavior is illustrated according to the driver behavior measurement method that was used for the analysis in the previous studies. The discussed measurement methods included simulation environment studies, test tracks and on-road experiments studies, smartphone studies, and Naturalistic Driving Studies (NDS).

Chapter Three explains and compares the common driver behavior measurement, including driving simulators, test tracks, on-road experiments, and NDS. Moreover, this chapter discusses the reasons for using the NDS data in this dissertation and, briefly, presents how this data was used in the literature. Then, the data is described, and the procedure of extracting and mapping the events of interest is explained. Finally, this chapter discusses the results of quantifying the behavioral measures for the extracted events.

Chapter four proposes an additional layer to the context-aware DAS architecture (i.e., context identification layer), which consists of two algorithms. The first algorithm, the infrastructure detection algorithm, is discussed in detail as well as how the neural networks are used in developing the algorithm. Moreover, the selection of the optimum neural network size is investigated. Finally, this chapter analyses the results of the used neural network models to study the influence of input variables on the models' accuracy.

Chapter five discusses the second algorithm, the driver classification algorithm, which is part of the proposed context identification layer. This chapter first defines the driver behavior classes used by previous studies and outlines the classes used in this dissertation. Then, the location and driver behavioral classification procedures are illustrated. The differences between the results of these classifications are examined. The location-based behavioral classification is highlighted as the proposed driver classification algorithm.

Chapter six summarizes the conclusions, findings, contribution of this dissertation, and potential future research areas.

2. LITERATURE REVIEW

This chapter provides a review of how driver behavior is integrated into the development of context-aware DAS. Previous studies that addressed driver car-following behavior during braking will also be discussed. Finally, this chapter discusses the classification as well as the identification and detection of aggressive driving as a mean to individualize the context-aware DAS. The discussion around drivers' classification will be split based on the data sources, i.e., driving simulators, test tracks, and on-road experiments, smartphones, and naturalistic driving data.

2.1 CONTEXT-AWARE DAS

Context-aware DASs have been adopted in several transportation applications. For instance, these applications could help in learn the transportation mode using smartphone data (Fang et al., 2017). For parking, such applications were used to recommend a nearby spot for a driver using an onboard unit with parking preferences (Alhammad et al., 2012). For cyclists, a context-aware application was proposed to assess the risk of cyclists' routes using video images and smartphone sensor data (Costa et al., 2017).

In addition, context-aware systems were used to detect various aspects of driver behavior. For example, Al-Sultan et al. (2013) suggested a context-aware system to detect driving behavior (e.g., normal, drunk, fatigue, and reckless) in real-time in Vehicular ad hoc Networks (VANETs). Contextual information about the driver, the vehicle, and the environment was used for the detection. This information was split into measures that affect behavior (e.g., human sleep-awake cycle and driving environment) and measures that result from specific behavior (e.g., speed limit violation, lateral position maintenance, acceleration, eyes' movement, and Alcohol intoxication).

The environmental factors defined to cause fatigue are temperature and noise, regardless of drivers' location (Al-Sultan et al., 2013).

Böhmländer et al. (2017) proposed a five-layer context-aware system to predict potential vehicle-to-vehicle head-on collisions, by triggering the irreversible safety actuator before collision. The reasoning layer in this architecture integrated several models and algorithms. The first algorithm is the environment sensing and the object tracking algorithm to estimate the kinematic characteristics of the host vehicle, and track and identify the surrounding objects (using video images and a group of sensors). The unavoidable crash was predicted using a collision risk model based on the trajectories of the host vehicle and the surrounding vehicles, and collision severity probabilities. These probabilities were anticipated using two physical models to estimate the vehicle mass and frontal vehicle stiffness. Finally, using the outputs of the models, a crash severity data model was used to define an activation threshold to trigger the safety actuators (e.g., smart airbags) prior to a collision (Böhmländer et al., 2017).

Using smartphone data, Bejani and Ghatee (2018) proposed a context-aware system that consists of four subsystems to detect dangerous drivers by assessing driver style. The context was defined within the context awareness subsystem by the traffic condition, maneuver type, driver style (i.e., dangerous or normal driver), and car type. Traffic conditions were limited to two levels, namely, congested and normal. The levels were detected using neural networks based on the GPS speed data, which showed enough accuracy to differentiate between congested and normal conditions. For the maneuver type classification, U-turn, turn, lane change, and no maneuver were the four main actions the system could classify using a hybrid machine learning technique (i.e., includes decision tree and neural network). The car type classification refers to the extent the car could affect the acceleration sensed inside the vehicle cabin. The proposed system uses a wide range of machine learning algorithms, including but not limited to neural networks, decision trees, *K*-means, and rule-based fuzzy algorithm. The system was evaluated by volunteer drivers in an on-road experiment setting, and it showed highly acceptable accuracy. Nevertheless, the classifications provided by the system were either high level (e.g., traffic condition level classification) or limited to specific types (e.g., driver style and maneuver classification) (Bejani and Ghatee, 2018).

Gupta et al. (2014) proposed a context-aware architecture that using wireless sensor network in VANET context, could detect five different driving styles, namely, drunk, fatigue, rash, wayward, and acceptable driving. This architecture was built to gather context information from GPS, cameras, and several sensors (e.g., an Alcohol sensor, speed sensor, accelerometer sensor, and surround sensor). The collected information was processed, and the state of the driver was inferred using swarm intelligence algorithm. It is worth mentioning that the proposed architecture was presented in a conceptual design format (i.e., the architecture was not tested) (Gupta et al., 2014).

Sathyanarayana et al. (2011) considered drivers' maneuvers as a component for a context-aware system that was proposed to work with active vehicle safety systems (e.g., adaptive cruise control, lane departure warning, etc.). The recognized maneuvers were left turn, right turn, and lane change in both residential and commercial areas. Nonetheless, these maneuvers were not investigated according to the location. It is noteworthy that the main purpose of the proposed system was to detect drivers' distraction through three sub-tasks, namely, driver identification, maneuver recognition, and distraction detection. Also, to detect driver's style and to recognize his/her maneuver, mobile phone input methods (i.e., GPS, rear-facing camera, accelerometer, and gyroscope) were used to collect contextual data about the driver and the vehicle kinematics in urban, rural, and highway environments and to detect aggressive turning maneuvers (Johnson and Trivedi, 2011).

Satzoda and Trivedi (2015) proposed a context-aware lane estimation framework to support lane changing and lane keeping DAS in terms of enhancing the computational efficiency of lane detection from video inputs without affecting the detection accuracy. However, the main focus of this study was on lane changing applications (i.e., lane estimation for lane changing DAS), which requires lane estimation in a far depth of vehicle's sensors view. The proposed framework consisted of two main functional modules, namely, the context definition module and the lane detection module. The proposed framework combined vehicle dynamics, surrounding information, and system-level requirements (i.e., application and accuracy requirements) to define the context in the first module (i.e., context definition module). The vehicle dynamics and surrounding information were collected using sensors installed in the vehicle beside the CAN bus data. This module was used to define how a lane detection module will operate in terms of how many of the input images should be processed based on the context. In brief, the context was defined based on the distance to the leading vehicle, the speed of the host vehicle to estimate the traffic condition, and the curvature, which was estimated as the ratio between the change in the angle of the vehicle motion vector (estimated from the yaw rate) and the distance traveled during this change (Satzoda and Trivedi, 2015).

Chang et al. (2017) proposed a three-layer context-aware service system to provide the driver with a customized business service (e.g., route selection service, gas station selection service, hospital selection service, etc.) according to the driver, vehicle, and environmental statuses. The three layers were the context acquisition, data transmission, and business service provision. The context acquisition layer gathers the contextual information (i.e., driver, vehicle, and environment-related data) from various sources. This information included driver's vital signs (e.g., blood pressure, heartbeat, etc.), vehicle-related information (e.g., speed, gas volume, in-vehicle temperature, etc.), and surrounding environment-related information (e.g., weather information, road accidents, etc.). The data transmission layer is responsible for transmitting the data to a management center, other vehicles, or roadside equipment through VANET or mobile ad hoc network (MANET). The last layer, business service provision layer, contains the context reasoning layer, which interprets the transmitted data, identifies abnormal situations, and recommends a business service to the driver. The recommended business service is proposed to mitigate a potential hazard situation, such as choosing a nearby hospital, to a driver who is in a critical situation (Chang et al., 2017).

Wan et al. (2014) proposed and discussed a multi-layer context-aware DAS architecture that enables the cloud capability to provide reliable and robust service for the users. The proposed architecture combined vehicle social networks and vehicular security to the context-aware DAS. This combination was obtained by adding a location computational layer and a cloud computational layer to the architecture. The location computational layer depends on the DSRC between in-vehicle OBEs and RSUs to accurately locate the vehicle. The cloud computational layer includes multiple cloud systems from various stakeholders (e.g., traffic authorities) to share relevant information and resources. The proposed architecture was exemplified by presenting a context-aware dynamic parking service (Wan et al., 2014).

Yang et al. (2015) proposed a context-aware DAS based on the history of driver's trajectories to select best route option according to driver's preferences. The context was defined in terms of trip timing (i.e., morning, afternoon, etc.) or traffic flow conditions (i.e., peak or off-peak). The context-aware driver's preferences for each context were identified in this study based on a combination of travel distance as static travel cost (measured from the maps), and travel time and fuel consumption as dynamic travel costs. The travel time and fuel consumption (calculated based on the average speed and acceleration) were measured based on the collected historical GPS trajectories. The proposed methodology was supported by a case study that collected GPS records from several drivers in Denmark (Yang et al., 2015).

For fuel-efficient driving, Gilman et al. (2015) suggested a context-aware DAS architecture that focused on the individualization and the adaptability of the system. The proposed architecture was presented as an "Action-Perception loop" system that gathers and detects relevant information and decides on actions that could change the driving context (e.g., driver behavior). This change, which will be detected by the system, will require new actions and so on. This system was developed by adding a meta-level layer that monitors and evaluates the reasoning and the decision-making process in the context-aware DAS. An application called "Driving coach" was presented in this study as an example of the proposed architecture. However, this application was a prototype and did not implement all the components of the proposed architecture. For example, this application was not a real-time feedback application meaning that it gave web-based feedback to the driver after the trip. The main purpose was to provide insights into fuel consumption during the trip. These insights were described based on performance indicators (i.e., thresholds for good and bad trips in terms of fuel consumption) that came from statistical distribution of related factors (e.g., speed distribution on different bins, acceleration, deceleration, and engine speed). These distributions were generated based on the driving history, which was over two weeks, and could be adjusted according to the weather as an environmental component of the driving context (Gilman et al., 2015).

Moreover, the performance and the acceptance of a context-aware intelligent speed adaptation system was assessed using data from several taxi drivers (Hoogendoorn et al., 2012). It was found that the system could reduce fuel consumption and acceleration significantly. However, the acceptance of the system was relatively low because the system did not consider the surrounding traffic and defined the surrounding environment based on the speed limits only (Hoogendoorn et al., 2012). Another context-aware system was used to optimize an energy-driven route for electric vehicles (Y. Wang et al., 2013). This system considered real-time traffic data to estimate the routing costs (e.g., travel time and energy consumption) of taking alternative routes. The real-time traffic conditions and the energy cost function on road links are estimated mainly based on the average speed, and samples of instantaneous speed and its derivatives (i.e., acceleration/deceleration). The system is expected to guide the vehicle to a defined destination through an optimal energy-driver route or to a recharge point if the defined destination was not reachable (Y. Wang et al., 2013).

2.2 Car-Following Behavior During Braking

Generally, car-following behavior is an extensive area of research due to the high frequency of this particular behavior. Previous studies have presented insights into this behavior that were beneficial for various purposes, such as car-following models' development and calibration (Hammit et al., 2018; Rakha, 2009; Rakha and Crowther, 2003; Tang et al., 2012; Treiber et al., 2000), rear-end near-crashes and crashes investigation (Arbabzadeh and Jafari, 2018; Carney et al., 2016; Gettman et al., 2008; Tawfeek and El-Basyouny, 2018a; Wu and Thor, 2015; Yang Zheng et al., 2014), and DAS development (Barber and Clarke, 1998; Bella and Russo, 2011; Doi et al., 1994; Fujita et al., 1995; Hirst and Graham, 1997; Kiefer et al., 1999; Tawfeek and El-Basyouny, 2018b).

These studies include several aspects of car-following behavior. One of the crucial car-following behavior aspects is the braking maneuver, since braking was indicated as the first response by the following vehicle driver to leading vehicle's deceleration, hard brake, or stopping (Adams, 1994; Lee et al., 2007).

The discussion in this section is split into three stages, namely, pre-braking, during braking, and post-braking to summarize the findings in the literature. The pre-braking stage includes the driver's intention to brake and the start of the braking maneuver (i.e., the perception, decision, and the start of the action). Once the driver starts to brake, the pre-braking stage ends and the during braking stage starts. During this stage, the driver might prepare to stop or slow down for different reasons, such as approaching a controlled intersection or approaching an end of a queue. Moreover, the driver might brake to avoid a near-crash or a crash. The during braking stage ends when the driver reaches a stopping condition or moves his or her foot from the brake pedal. On the other hand, the post-braking stage defined in this dissertation occurs when a braking event ends with a situation where a near-crash or a crash occurred after an evasive maneuver, such as hard braking, or sudden steering, or moving the foot from the brake pedal, or accelerating in other safe situations.

For the pre-braking stage, Wang et al. (2018) proposed a Gaussian mixture model with a hidden Markov model to predict the driver braking in car-following situations. The model prediction of the braking action was in the form of a binary output where "1" means the driver will use the brakes. This model was trained using vehicle speed, time-to-collision, relative speed, and range (i.e., the distance between the host vehicle and the leading vehicle) as independent variables to describe the car-following driving state. These variables were extracted from NDS data for forty-nine drivers. The trained model was compared with two support vector machine based approaches (i.e., basic support vector machine and support vector machine with Bayesian filtering) (Wang et al., 2018).

Moreover, Bella and Russo (2011) studied driver's behavior before taking certain maneuvers in a car-following situation (i.e., overtake maneuvers or braking actions) to develop an FCW algorithm. In their experiment, Bella and Russo (2011) used a driving simulator with 32 drivers that generated a total of 315 lane-changing (i.e., overtaking) maneuvers and 160 car-following events. These events were extracted based on a definition of an acceptable alert timing zone for the FCW around 1.18 to 1.52 seconds before a driver's maneuver to avoid a rear-end incident. Similar thresholds for alert times were suggested by Kiefer et al. (1999). Using linear regression and based on the extracted events, an FCW algorithm was developed (Bella and Russo, 2011).

Using a small sample of NDS and with the aid of the in-vehicle video recordings, Chatterjee and Davis (2014) aimed at analyzing drivers' behavior on congested freeways as a preliminary step to characterize their behavior in a freeway shockwave. For this analysis, a methodology was proposed to detect a window of the braking events during a car-following situation on the congested freeway. These were described as the events in which the following vehicle brakes at a higher rate than the leading vehicle. This methodology identified relevant events based on several parameters such as initial leading and following vehicles speeds, the distance between the leading and the following vehicle, and, as assumed, deceleration rate. The results showed that 14 out of 15 braking events were successfully identified. It is worth mentioning that the methodology was verified using some video recordings and a few false positives were detected from the methodology. However, these false positives were not discussed (Chatterjee and Davis, 2014).

For the during braking stage, several behavioral measures were investigated. For example, In a non-car-following situation, Casucci et al. (2010) defined the braking behavior as the actual use of the braking pedal and measured by the maximum force applied on the brake pedal and the maximum brake pedal usage duration. The main aim of this study was to present and describe a simulation tool that integrates a set of algorithms that model driver-vehicle-environment interactions in a computer program. The results of the described simulation tool were compared to the outputs of a virtual reality driving simulator to validate the simulation tool. The simulation tool was validated with two maneuvers, namely, braking behavior and lateral performance. In the driving simulator, the braking behavior was tested in relation to three other variables, including driver's experience, hazardous events on the road (e.g., pedestrian crossing, a vehicle suddenly turning out in front of the driver and a stopped vehicle on the side of the road), and distraction level. For the braking behavior, two main hypotheses were tested to differentiate between the drivers based on their experience and based on their distraction level. The results showed that the driving experience did not have a significant impact on the braking behavior in terms of the two used measures. On the other hand, distraction had a significant impact on the braking behavior (Casucci et al., 2010).

For car-following situations and using the Naturalistic Driving Study (NDS) data to investigate rear-end events, Montgomery et al. (2014) studied time-to-collision (TTC) of following behavior while braking. The analysis included comparisons between differences in TTC for various ages and gender. It was found that there was a statistically significant difference between various demographics of drivers while braking. In addition, Chen et al. (2016) compared the distributions of TTC and enhanced TTC (ETTC) at braking events in normal car-following situations. ETTC is similar to TTC but considers the relative acceleration between the following and the leading vehicles. The analysis of the data showed that the ETTC has less variation between drivers with different speed bins.

The TTC and headway were compared as safety indicators in different locations around an intersection with a stop sign on the minor road (Vogel, 2003). TTC and headways during braking were collected using 7-point stations on main and minor roads. Two of these stations were setup near the intersection (17.5 meters away from the center of the intersection) on the main road with no stations near the intersection on the minor road. The results showed that there was no difference between the percentages of small headway values (less than 1 second) close and far away from the junction. On the other hand, the comparison of TTC values showed that one of the two stations near the intersection had the largest percentage of small TTC values (Vogel, 2003). It is worth mentioning that the results of this study were compared visually, and no statistical testing was conducted.

Bella et al. (2014) compared time headway and TTC on highways during different times of the day (i.e., day and night), following different vehicle types (i.e., passenger cars and heavy vehicles), and travelling lane type (i.e. left or right lane) to investigate the difference between the two measures. The results showed that time headway and TTC are independent, which indicates that both measures provide different insights about drivers' car-following behavior. Based on the results of this study, TTC is more effective than time headway to identify impending risk of a rear-end collision (Bella et al., 2014).

To evaluate safety, comfort, and trust of an FCW system, the visual input of the driver and his behavior when following a leading vehicle during braking were analyzed (Wada et al., 2007). Driver behavior was expressed in terms of the distance to the leading vehicle and "performance index for approach and alienation", which was proposed as the driver's visual input. This index is based on the area change of the leading vehicle on the retina, which is a function of gap distance and relative velocity. The relation between the proposed index and the gap distance were validated by a test driver a real driving trip and 4 drivers in a driving simulator. The results imply that the ratio between the proposed index and the gap could be used to assess the unsafe driving situation and develop driver assistant systems (Wada et al., 2007).

For the post-braking stage, the braking process of the vehicle and surrounding environment were investigated to study the factors affecting near-crashes using NDS data (Yang Zheng et al., 2014). In this study, near-crashes were identified and collected when the vehicle deceleration reached a certain threshold. Near-crash data was recorded every 10 seconds before and 5 seconds after the identified near-crash. Three driving risk levels (i.e., low, medium, or high potential for vehicle crash or other accidents) were represented in near-crash events by the maximum deceleration, average deceleration, and the loss in the kinematic energy from the braking trigger to maximum deceleration. The number of the identified near-crashes were distributed on 19 contributing factors. Vehicle speeds when braking starts and braking triggering factors were found to have the largest impact on driving risk levels. Braking trigger factors included non-host vehicle, traffic light change, lane reduction, lane change, and collision avoidance. Near-crash location was among the 19 factors studied in this research, and the results showed that near-crash location was not a significant contributing factor to the driving risk level. However, intersection near-crashes were higher from a driving risk level perspective when compared to non-intersection near-crashes (Yang Zheng et al., 2014).

Bagdadi (2013) assessed the use of the jerk (i.e., the rate of change in longitudinal acceleration) as a measure for critical safety braking detection as a post-braking stage. The jerk was proposed by (Bagdadi and Várhelyi, 2013) as a crash proneness indicator. The assessment was performed on predefined near-crash situations which were extracted from NDS data (Bagdadi, 2013). The jerk was also compared to longitudinal acceleration as critical safety braking detection measure. The results identified a certain threshold for critical jerk (i.e., 1.0 g/s) to detect critical events. Moreover, the comparison between the jerk and the longitudinal acceleration indicated that the jerk was substantially more efficient than the longitudinal acceleration for the detection of near-crashes (Bagdadi, 2013). It is worth mentioning that this methodology was applied only on a set of predefined near-crashes and was not applied to an entire NDS data. This might raise some concerns about the proposed threshold since it might not be accurate enough in identifying the near-crashes as highlighted by Perez et al. (2017). In this study, two NDS data sets were analyzed to develop several thresholds for some behavioral measures to identify near-crashes and crashes. Moreover, the proposed thresholds were validated based on false negative assessment analysis. However, the results showed that the events identified from the proposed thresholds were too many (i.e., over half a million), which highlighted the importance of improving the proposed methodology. This suggestion faces the complexity and the massive size of the NDS data which, makes the validation process infeasible (Perez et al., 2017).

2.3 The Detection of Drivers' Behavior Variation

The current literature includes several approaches to detect the variability of drivers' behavior using either questionnaires and surveys or various measurement methods (e.g., driving simulators, test tracks and on-road experiments, smartphones, and naturalistic driving data) or both. Since the main goal of this dissertation is to enhance the development of DAS and driver behavior questionnaires and surveys are not suitable (Wang et al., 2010b), this section will focus more on the studies that used the behavior measurement methods to detect variation. The discussion in this section will be split according to the measurement method used for studying the behavior variation.

2.3.1 Drivers' Variation in Simulation Environments

Rong et al. (2011) used a driving simulator and microsimulation software package (VISSIM) to investigate drivers' dissimilar behavior and the effects on traffic flow characteristics. The study combined a driving simulator to classify drivers and to obtain measures (i.e., reaction time, and expected speed) for the car-following model and a microsimulation model calibration. The driving simulator scenes were designed based on a mixture between hypothetical three multilane roadway and real roadside buildings and roadway design. The driver classification was based on 15 indicators, which were split into two groups, namely, physical and psychological indicators (e.g., age, driving experience, blood pressure, etc.), and driving behavior indicators (e.g., average and maximum speed, maximum acceleration, minimum gap acceptance, and lane change frequency). The Principal Component Analysis (PCA) was used to eliminate any possible correlation between the selected 15 indicators. This was achieved by extracting seven principal components based on the PCA as the main variables for the classification. The output of the classification was then used to study the effect of driver behavior on the traffic flow stability. The results indicated that, unlike conservative and moderate drivers, more aggressive drivers in the network could positively impact traffic flow but, on the other hand, reduce the traffic flow stability (Rong et al., 2011).

Again, using a driving simulator, a semi-supervised technique was proposed to classify drivers into aggressive and normal drivers with a small number of labeled data (Wang et al., 2017). Vehicle speed and throttle opening were recorded for 20 drivers from the simulator on a hypothetical closed curvy path to define aggressive driving without considering any surrounding traffic. Based on these two measures, *K*-means clustering was used to label a small portion of the data, and the semi-supervised support vector machine was then applied. The results showed that the accuracy of the proposed technique, which needed a small amount of labeled data, outperformed the supervised techniques (i.e., support vector machine with linear kernel) (Wang et al., 2017). Moreover, Casucci et al. (2010) differentiated between the prudent and aggressive drivers based on braking behavior and lateral performance. It was found that aggressive drivers completed an overtaking task faster than prudent drivers. In addition, the cruising speed (i.e., the final speed after completing the overtaking maneuver) of aggressive drivers is higher than for prudent drivers (Casucci et al., 2010).

Chandrasiri et al. (2012, 2016) categorized sixteen drivers into two classes according to their driving skill on curves (i.e., high, and low/average skilled drivers). Ten behavior attributes that represent lateral and longitudinal behavior (e.g., steering angle, longitudinal acceleration, and lateral acceleration) were collected using a driving simulator. Drivers classification was executed using two machine learning techniques, namely, K-nearest neighbor and support vector machine with different kernels. The selected machine learning techniques were then implemented on the features that were extracted from the collected data using Principal Component Analysis (PCA). It is worth mentioning that several runs were executed with a different number of principal components and the corresponding accuracy was estimated.

Yang et al., (2018) suggested two layers of car-following behavior recognition by monitoring the physiological signals of fifty-two drivers during six traffic states. Using a driving simulator, physiological signals and fourteen driving behavior measures (e.g., longitudinal and lateral acceleration, relative speed, space and time headway, etc.) were collected on a hypothetical two-way, two-lane straight road segment. Based on the experiment design, the drivers were not allowed to overtake the leading vehicle to enable testing six traffic states. These traffic states were simulated according to the leading vehicle speed profile. The first layer of the driver behavior recognition combined *K*-means clustering with support vector machine technique to classify the drivers to five categories. These categories encompassed five driving styles and skills, namely, normal, aggressive stable, aggressive unstable, unaggressive stable, unaggressive unstable driving. The second layer of the proposed method included the linear discriminant analysis to extract two core physiological signals, and this layer used the outputs of the first layers as inputs to train a *K*-

nearest-neighbor algorithm to classify the drivers. The average accuracy of the proposed method for the six traffic stated was 69.5% (Yang et al., 2018).

To assess how being late or being under time pressure influence driver aggression, Fitzpatrick et al. (2017) used a driving simulator with a hypothetical rural two-lane roadway to collect several behavioral measures including speed, acceleration, gap acceptance, passing a slow-moving vehicle, and running a yellow light. These behavioral measures were used to differentiate aggressive driving by comparing two experimental groups to a control group. These two experimental groups (i.e., hurried and very hurried drivers) were put under time pressure by offering the participants some incentives if they were able to finish the test within a specific time period. However, the simulated test route included several signalized intersections, but drivers' behavior near these intersections was not compared to their behavior elsewhere in the test route. The results indicated that the very hurried drivers, who were under high time pressure, were more likely to drive aggressively in terms of the above-mentioned measures when compared to the control group (Fitzpatrick et al., 2017).

2.3.2 Drivers' Variation in Test Tracks and On-Road Experiments

Sundbom et al. (2013) defined aggressive and normal driving based on steering behavior in terms of curve lateral acceleration, yaw rate, and lateral movement in the lane. Only two drivers participated in this study on a test track where each driver had a mixture of aggressive and normal driving. A model was developed to classify the two driving styles online (i.e., aggressive and normal). It is worth mentioning that the test track had no traffic and consisted only of curved sections and straight sections. Also, the straight segments were removed when estimating the model performance because of drivers' lack of excitation in these segments, which generated classification errors (Sundbom et al., 2013).

In an on-road experiment, the dissimilarity between drivers was studied using measurable parameters that were related to time headway, the inverse of TTC, and the time between accelerator pedal release to brake activation (Wang et al., 2010a). The drivers were classified using *K*-means clustering according to their aggression, stability, conflict proneness, and skill. The results of this study showed that the *K*-means clustering technique was efficient in detecting the dissimilarity in the longitudinal driving behavior.

Using speed and lateral and longitudinal acceleration, Gonzalez et al. (2014) developed a linear filter based on a Gaussian Mixture Model to detect driving aggressiveness. This model was validated by monitoring the behavioral measures of 10 drivers on a customized test track. This test route, which was part of a private road network, included five different road features such as a roundabout, curves, uphill and downhill segments, and a sharp bend. The drivers conducted several driving rounds on the route including only one aggressive round for each driver. It is worth mentioning that the drivers were not given any instruction or definition of aggressive driving (Gonzalez et al., 2014).

Zheng et al. (2017) defined aggressive driving based on a questionnaire given to 15 drivers before and after local on-road tests (within the University of Florida campus area). This study aimed at assessing the relationship between the driver type and driver behavior in a high vehicle-pedestrian interaction low-speed environment. This aim was achieved by classifying the drivers into four types (i.e., type A, B, C, and D) with the corresponding aggression index, which was obtained from the questionnaire. The environment was on the campus, with speed limits between 20 mph and 35 mph. The driving aggressiveness obtained from the questionnaire was compared to driving the desired speed and yielding to pedestrians. The results revealed that the two observed measures could efficiently identify aggressive drivers (Zheng et al., 2017).

Hill et al., (2015) used two behavioral measures (i.e., discretionary lane changing and speed ratio) extracted from forty-six drivers during an on-road experiment on two different freeways to classify drivers based on their behavior. The data collected in this study was associated with a self-assessment driving habit questionnaire. The aim was to classify the participating drivers into four groups according to their actual behavior, namely, very conservative, somewhat conservative, somewhat aggressive, and very aggressive. This classification was performed using the *K*-means method based on the discretionary lane changing per kilometer and the speed ratio (i.e., the ratio between the free flow driving speed in a non-car-following situation and the posted speed limit). The preliminary analysis of the established measures (i.e., the frequency of discretionary lane changing and speed ratio) for the data of the two freeways showed that there was a difference between the behavioral measures of the two freeways. To address this, the clustering analysis was applied on each freeway separately, and then the behavioral trends in each of the four groups were analyzed to highlight the importance of considering drivers' variation within the behavioral models (Hill et al., 2015).

2.3.3 Drivers' Variation using Smartphones

Based on smartphone accelerometer data (i.e., the acceleration in X, Y, and Z axes of the smartphone), Osafune et al. (2017) classified the drivers as safe or risky. This classification was validated based on drivers' crash records over 20 years. The driving data, including GPS positions and acceleration was collected using a smartphone application for 15 months, covering almost all areas of Japan on several road types (urban, suburban, and highways). Thresholds for three behavioral measures, namely, acceleration, deceleration, and left acceleration were proposed as accident risk indexes for identifying risky drivers. These accident risk indexes were then verified using the one, two, and three-dimensional support vector machine. The results showed that the proposed risk indexes could successfully classify the drivers as risky or safe with approximately 71% accuracy (Osafune et al., 2017).

To detect aggressive actions using a smartphone without external processing, smartphone data from three different drivers in urban, rural, and highway environments were used (Johnson and Trivedi, 2011). The data was collected from smartphone's rear-camera, accelerometer, and GPS. Based on the Dynamic Time Warping algorithm, the system was suggested to detect and classify aggressive or normal turning and lane change maneuvers. However, the system identified the aggressive turning movements only at U-turns (Johnson and Trivedi, 2011).

Similarly, to detect risky driving and classify drivers, a proposed smartphone application was created by fusing GPS data obtained from smartphones (e.g., speed variation, and bearing variation) with smartphone sensors data (e.g., jerk standard deviation, and average yaw rate) (Castignani et al., 2015). Along a predefined path, the participating drivers were instructed to drive calmly (i.e., observing the speed limit and avoid sudden maneuvers) in the first lap and to drive aggressively in the second lap. All the collected measures were reduced to two principal components and used to cluster the drivers into aggressive and calm. It is noteworthy that the distributions of aggressive events were clustered at intersection areas, which indicated that drivers are more aggressive at intersections. However, this observation was drawn without a specific definition of the intersection area and without considering the interactions with surrounding vehicles (Castignani et al., 2015).

Based on smartphone sensors, Singh et al. (2017) proposed a real-time system that could detect aggressive driving using Dynamic Time Warping technique. Aggressive driving was defined in this study as sudden braking events in lower speeds (i.e., between 20 and 30 kph). These braking events may reflect congestion when occurring over a short time span or repeated from several vehicles in the same area. Eftekhari and Ghatee (2018) also proposed a system that is a hybrid of Discrete Wavelet Transformation and Adaptive Neuro-Fuzzy Inference System to classify the drivers into safe, semi-aggressive, and aggressive drivers. Drivers' self-reported questionnaire was used to train and test the proposed system. The results showed that the system could detect drivers' dissimilarities with an accuracy of 92%. Moreover, a sensitivity analysis was conducted on the measures extracted from the smartphone sensors, and the results indicated that the longitudinal acceleration is essential for driver behavior classification. Conversely, drivers still could be classified with sufficient accuracy without the angular velocity extracted from the smartphone sensors (Eftekhari and Ghatee, 2018).

Bejani and Ghatee (2018) proposed a four subsystems context-aware system to evaluate driver style. One of these four subsystems was a context awareness subsystem, which included the driver style classification. The system depended on classifying drivers as dangerous or normal for only three maneuvers, namely, lane change, U-turn, and turn maneuvers. To classify the driver style, ten acceleration inputs were used. These acceleration inputs included the average and the variance of the acceleration on the longitudinal and lateral directions before, during, and after the maneuver. A combination of *K-NN*, support vector machine, and multi-layer neural network were fused to detect dangerous driving. The proposed system was validated in an experiment that involved 27 drivers travelling a specific route with two experts as passengers to evaluate the drivers' style (Bejani and Ghatee, 2018).

2.3.4 Drivers' Variation in Naturalistic Driving Environments

Feng et al. (2017) assessed the longitudinal vehicle jerk as a measure to detect the variation in drivers' behavior using NDS data. This detection was mainly done by identifying aggressive drivers. Hence, two jerk-based measures were proposed, namely, the frequent utilization of large positive or negative jerk. First, drivers were recognized as aggressive according to their excessive speeding, tailgating, crash or near-crash frequency in the NDS data. Based on that, three groups of aggressive drivers were generated, and the proposed jerk-based measures were validated using these three groups. The results showed that the proposed measures could successfully identify aggressive drivers in the three groups. However, the results indicated that the large negative jerk was more efficient when compared to the large positive jerk in detecting aggressive drivers (Feng et al., 2017).

Based on a subset of NDS data, Arbabzadeh and Jafari (2018) proposed an approach to predict the driving outcomes, which were described as five driving states: normal driving, which had three types; near-crash; and crash. This study highlighted the drivers' variation by defining a driving status spectrum that starts with safe driving and gradually the risk increases to a fatal collision. The start of the driving state spectrum and the first type of the normal driving state was defined as the safest driving state. This type did not include any unsafe driving behavior or any involvement in a non-driving secondary task. The second type of the normal driving state involved the engagement in at least one secondary task without any unsafe driving behavior detected. Finally, the third type of normal driving state included undertaking at least one unsafe driving behavior attribute or maneuver and the involvement in at least one non-driving secondary task. It is worth mentioning the top unsafe driving behavior attribute or maneuver included excessive speeding, fatigue or drowsiness, failure to signal, stop sign violation, and driving slower than the surrounding traffic. Moreover, the top secondary tasks included interaction with a passenger in the adjacent seat, talking to the unknown audience, talking, holding, texting using a cell phone, and adjusting the radio. These unsafe driving behavior attributes or maneuvers and the secondary tasks were already reported in the NDS data.

For freight transport industries, Wu et al. (2016) proposed a method to classify commercial truck drivers according to their behavior. The proposed methodology was applied to GPS data for fifty commercial medium-sized truck drivers. Eight parameters related to speed and acceleration or deceleration were used in this study. These parameters were both the average and the standard deviation of the speed, the deceleration and the acceleration, and the proportion of the time spent driving above 80% of the speed limit. The PCA was used to represent most of the observed variance in the dataset into a smaller number of factors (two principal components). Hierarchical clustering analysis was then conducted to define five levels (i.e., minimal, slight, moderate, serious, and severe) of drivers' acceleration, deceleration, and speeding prone behaviors. The drivers, who were identified to have severe behavior in terms of acceleration., deceleration, and speeding, were judged as risky drivers (Wu et al., 2016).

Gilman et al. (2015) collected driving behavior data for a fuel-efficient context-aware application. The proposed system architecture included behavioral measure thresholds to identify aggressive driving. Although the proposed system collected various behavioral measures, those used to identify aggressive driving were acceleration, deceleration, and speed. Moreover, aggressive driving detection was not fully functioning in the proposed application (Gilman et al., 2015).

The NDS data was also used to detect variations among older drivers (Guo et al., 2015). This study aims at assessing older drivers' fitness to drive using naturalistic driving data of 50 older drivers. This assessment was carried out by studying the relation between older drivers' fitness profile and the risk of being involved in a crash. The PCA was used to extract non-correlated features (i.e., principal components) from 53 fitness-to-drive metrics. These metrics were grouped according to the nature of the metric into four groups, namely, health, visual, physical, and cognitive ability. For each of these four groups, a set of principal components (i.e., 16 principal components) were generated and investigated through negative binomial regression analysis. The regression analysis was used to develop the link between the crash and near-crash data and the drivers' fitness profile. The results showed that the right eye contrast sensitivity metric (i.e., the ability to distinguish between visual targets such as pedestrians that may have low contrast with the background) was significantly affecting the crash and near-crash risk (Guo et al., 2015).

Drivers variation was investigated by recognizing drivers' pattern in a subset of NDS, which included ten car drivers and ten truck drivers (Higgs and Abbas, 2015). The pattern recognition process split into three main steps, specifically, car-following events extraction, event segmentation, and clustering. Eight behavioral measures were used in this pattern recognition process. These measures are longitudinal and lateral acceleration, yaw rate, range (i.e., the distance between the following and leading vehicles), range rate (i.e., relative speed), yaw angle, the following vehicle speed, and lane offset (i.e., the distance between the travel lane centerline and the centerline of the vehicle). Through the clustering analysis, the car-following events of each driver were split into thirty clusters. However, these thirty clusters were not attributed to specific behavior (i.e., braking, overtaking, merging, etc.). The car-following events were distributed on the thirty clusters indicating that the car drivers had a high heterogeneity in their behavior. In contrast, the distribution of the car-following events of the truck drivers was on specific clusters within the thirty clusters showing that the car-following behavior of those drivers is more consistent and homogeneous. It is worth mentioning that the output of the segmentation and clustering process was then used to calibrate a car-following model. The calibration results indicated that considering the segmentation and clustering process improved the performance of the car-following model. The summary of the findings of this study highlighted the importance of individualization when dealing with driver behavior (Higgs and Abbas, 2015).

Using a fleet of ten cars, Controller Area Network (CAN) bus data from fifty-four drivers were collected to highlight the differences between drivers (Fugiglando et al., 2018). The CAN bus signals used in the analysis included speed, longitudinal and lateral acceleration, engine's RPM, gas pedal position, brake pedal pressure, and steering wheel angle and momentum. Based on the collected data, a methodology was proposed to classify the drivers using *K*-means clustering. Although the data was collected in an uncontrolled environment similar to the NDS, the drivers did not use their own cars. This may have an influence on the participating drivers' behavior but collecting the data in an uncontrolled environment without route restriction made this study closer to naturalistic driving studies. Moreover, the CAN bus data was not collected with the corresponding drivers' location (Fugiglando et al., 2018).

2.4 SUMMARY

Although the idea of context-aware systems was first introduced in the early 1990s, there has been a great deal of recent attention focused on the development of context-aware DAS in the literature especially with the latest revolution in ITS. ***The first section*** in this chapter started with reviewing the literature by discussing the various context-aware DAS services and how these systems provided the proposed services. In general, the current context-aware DAS applications focused on providing the users (i.e., the drivers) with services that enhance the driving experience while making the driving process more efficient.

These services ranged from informative applications that provide the driver with optional information, such as guidance to available parking spots, a nearby gas station, to driving-related applications that require the driver's attention and action about safety critical situations such as FCW and LCW. Moreover, these services included advisories about reducing fuel consumption, minimizing trip travel time, or select the best travel route based on the driver preferences. In a

VANET or CV environment, the context-aware DAS could allow drivers to be notified about risky, fatigued, or impaired driving in their vicinity. Furthermore, the context-aware DAS was used to detect distracted and dangerous drivers, give the driver post-trip feedback about their driving, and saving energy for electric cars.

Despite this wide range of context-aware DAS, they either focused on the host vehicle driving data only or the host and surrounding vehicles driving data without considering the environmental component. Moreover, some of these systems defined the environment based only on the noise and temperature regardless of the driver's location. Some other systems studied driver behavior to detect the distracted or risky and aggressive drivers regardless of behavior change due to their location on the road network.

The second section of this chapter focused on studies of driver's behavior as a base for the development of context-aware DAS. Since car-following is the most frequent situation during the driving process and these situations reflect drivers' interaction with surrounding vehicles as part of the driving context, car-following was the main situation of interest in this section. Moreover, the braking maneuver was identified as the first normal response to deceleration or stopping by leading vehicle. To narrow the scope of the section, the car-following during braking events were the main focus. However, this section defined and addressed three braking stages, namely, pre-, during, and post-braking.

The pre-braking stage included the driver's perception about a need to brake either in a critical or normal situation. The research in this area discussed the prediction of the driver's braking action, the development of FCW based on drivers braking behavior, and investigating the driver braking response in traffic shockwaves. On the other hand, the post-braking stage was the end of the braking action when the driver removes his/her foot from the brake pedal or the braking ends with a near-crash or crash. Studies that tackled this braking stage involved the investigation of the contributing factors to near-crashes and crashes, proposing measures for crash proneness detection, or identifying thresholds to define near-crashes and crashes. For the during braking stage in car-following situations, the literature addressed studying the variation in the driver behavior based on the age and gender, comparing between different behavioral measures such as TTC, headway, and ETTC, or to evaluate FCW based on drivers' comfort and trust.

Although a considerable number of studies discussed and modeled car-following behavior, car-following while braking, specifically, did not get the same attention. Moreover, these situations (e.g., car-following during braking) were not quantified using behavioral measures other than the headway and TTC. Furthermore, most of these studies highlighted some insights about these situations (e.g., the ETTC had better results than the TTC to describe these situations) and recommended the consideration of such insights in the development of DAS but without discussing how this integration should be conducted. Also, the studies found in the literature did not address the effects of the surrounding infrastructure (e.g., intersections) on the driver behavior.

The third section in the literature review presented studies that investigated the classification of drivers based on their behavior or the detection of the variation among the drivers. The review was split based on the method that was used for the classification including driving simulators, test tracks and on-road experiments, smartphones, and NDS. Various statistical and machine learning techniques were used in the literature to classify the drivers into different groups and to distinguish between cautious and conservative, normal, and aggressive drivers. More discussion about the different classes of the drivers and the definition of each class will be presented in Chapter Five. The machine learning techniques, which were used to classify the drivers, included Gaussian Mixture Models, support vector machine, principal component analysis, and *K*-means. All these techniques were unsupervised machine learning techniques, which mean it was hard to judge the accuracy. However, these studies used to collect other information, which could support the explanation of drivers' classes such as questionnaires, drivers' age, driving experience, gender, being late, etc.

The driver classification studies used a wide range of measures for the classification. These included behavioral measures, physiological signals, fitness-to-drive metrics, or vital signs (e.g., blood pressure). For the smartphones, the accelerometer sensor data was used as the primary source for the behavioral measures, which were limited to speed and acceleration. For the other methods, various behavioral measures such as TTC, headway, yaw rate, and jerk or a combination of these measures were used to classify the drivers. Nevertheless, some of the studies focused on using one or two of these measures only. Some other studies used more than two measures and used a data reduction technique such as principal component analysis to abstract the data into two or more principal components. However, no studies were found that focused on combining behavioral measures (e.g., TTC, headway, following distance, acceleration, etc.) only to classify the drivers. The difference between the behavioral measures and other measures (e.g., physiological signals, fitness-to-drive metrics, or vital signs) is that the behavioral measures could be easily collected and processed in real time.

Despite all these efforts in the literature, the location factor was not considered while performing the classification or the detection of the aggressive driving. Thus, there was no link established in the literature between the aggregate behavior of the drivers and their behavior in certain locations on the network. However, some studies analyzed driver behavior and aggressive driving on only certain sections such as curves, yet these studies did not provide a link between the behavior on curves and the driver behavior in general. In other words, the drivers may tend to be more cautious or more aggressive on curves than normal due to the special design of these parts of the segments.

3. EVENTS EXTRACTION AND BEHAVIORAL MEASURES QUANTIFICATION

This chapter will discuss the widely used driver behavior measurement methods including driving simulators, test tracks, and on-road experiments, and the reason behind using NDS data for this dissertation. After describing the NDS dataset used in this dissertation, the car-following events and the corresponding behavioral measures extraction will be explained. Moreover, this chapter will discuss the effort of mapping and classifying the extracted events. Finally, the quantification of the behavioral measures will be presented in the form of probability distribution functions.

3.1 DRIVER BEHAVIOR MEASUREMENT METHODS

Several datasets have been used to study driver behavior and performance. These datasets provide a wide range of driver measurements and can be obtained from driving simulators, controlled test tracks, on-road experiments, and NDS. The selection of a specific measurement method depends on the purpose of the study and which element of the driver behavior will be investigated. Moreover, the literature includes some overlap between the definitions of some of these methods. For instance, the definition between the test track studies could overlap with the definition of the on-road experiments, and the same overlap could be found between the on-road experiments and NDS. Therefore, this section will discuss in detail each of the above-mentioned methods as well as the advantages and the disadvantages of each method. At the end of this section, NDS will be highlighted as the appropriate behavior measurement method for this dissertation.

There are different types of driving simulators; starting from desktop to multi-axis motion driving simulators with 360 degrees of display (Mclaughlin et al., 2009). Despite the flexibility they offer for testing several scenarios in a controlled environment, simulation results are often criticized for their limited ability to replicate real driving situations.

For this reason, several disadvantages have been reported about generalizing the findings from driving simulators, including lack of skill transferability. This phenomenon is related to whether or not the skills drivers learn in a simulated environment are transferred to the road environment. Transferability issues have a direct effect on the validity of simulations, which is not adequately addressed in the literature (De Winter et al., 2012). In addition, simulators discomfort and sickness are challenges that may affect drivers behavior and increase the probability of excluding drivers from the experiment (Cobb et al., 1999). Moreover, the fact that participants know they are not in a real situation might make them less cautious since there is no real risk of being involved in a collision (Evans, 1991). In addition, to avoid drivers' simulator-sickness, the study duration might be limited to a certain testing period (e.g., 10 minutes driving per session) and certain geometric features (De Winter et al., 2012). Even with the introduction of advanced driving simulators, this simulator-sickness cannot be eliminated entirely (Chen et al., 2018; Wang et al., 2016).

Certainly, controlled test tracks overcome the majority of the shortcomings associated with being in a driving simulator. Despite adding realism to the experiment, the driver is still considered in a testing environment which might make them behave in a way that is not consistent with how they would normally drive. For instance, controlled test tracks studies still share some limitations with the driving simulators such as the absence of the driver's destination. Because of that, the driver will not have the pressure that is generated from the motive of arriving at a specific destination (Smiley, 1996). In addition, the driver interaction with the surrounding vehicles and infrastructure is restricted to the test instructions either for safety purposes or for testing purposes (Smiley 1996, Gonzalez et al. 2014). These instructions range from detailed instructions such as driving at a certain speed (Boda et al., 2018) to general instructions such as driving aggressively (Gonzalez et al. 2014). Such instructions result in drivers deviating from their normal behavior and subsequently biasing the results of the analysis. In addition, when controlled test tracks are used to study drivers' response to imminent crash such as rear-end crash, a dummy rear-half of a passenger car could be used for safety (Kiefer et al., 1999). Similar to the driving simulator, the driver at this test setting is not driving in reality, which might affect their driving behavior (Kim et al., 2003).

To avoid the above-mentioned issues encountered with driving simulators and controlled test tracks, on-road experiments are also used extensively as a trial to capture the normal daily driver behaviour in the literature. On-road experiments are usually used in less hazardous situations which have minimal risk impact on the drivers participating in the experiment (Mclaughlin et al., 2009). During these experiments, the driver is fully aware of the driving situation with its typical risks that could occur. Therefore, the driver has that belief of reality which was absent in most of the cases in the other methods (i.e., driving simulators and test tracks). However, the drivers still could have the feeling of being in a test environment or being monitored. For instance, the driver could have a passenger from the research team in the vehicle as an observer. The role of the observer could be limited to observing the drivers' reactions and responses and take some notes. On the other hand, the observer could have an additional role during the experiment, which is the engagement in hazardous situations and take over some driving tasks from the main driver. This tasks could be, for example, the longitudinal control of the vehicle using a dual control pedal system (Naujoks et al., 2016b). Furthermore, the drivers are usually instructed to take a certain action or to drive on a certain road, which will increase the feeling of being in a test environment (Hill et al., 2015; Naujoks et al., 2016b). Moreover, the participants are not using their own vehicle during the on-road experiment. The driving data is usually collected using an instrumented vehicle that has special equipment, such as video cameras for data collection, depending on the experiment needs (Hill et al., 2015; Wang et al., 2010a) or a dual control pedal system for participant safety (Naujoks et al., 2016b). Driving a vehicle other than the driver's own could affect his or her typical daily behavior, especially with the instrumentation installed in the vehicle.

On the other hand, the NDS provides insights into normal daily driving without subjecting drivers to a test environment. NDS data provides a rich source of detailed, real-world, driving behavior data. NDS was conducted to collect normal everyday data about the driving context (i.e., the driver, the vehicle, and the surrounding environment) with hidden data collection equipment (Mclaughlin et al., 2009; UDRIVE, n.d.). Therefore, the NDS mitigates the disadvantages that encountered the driving simulators, the test tracks, and the on-road environment as driver measurement methods. For example, during the NDS, the drivers are not given any specific instructions and are driving their own vehicle on typical trips, which is not available in the other methods. This means that the drivers will have full control over not only the vehicle but also over their destination, which will be flexible according to their preferences and the route will not be restricted to the experiment route. This facilitates collecting the data over various highway facilities including freeways, arterials, collectors, etc., and over different traffic control devices including traffic signals, stop signs, yield signs, and so on. The variety and the flexibility offered by the NDS will allow a network-wide analysis of the collected data, and draw essential with respect to the network. Moreover, the drivers are collecting the data over a long time period (i.e., months) (Dingus et al., 2015, 2006). This longer data collection period will eliminate any potential test environment feeling that the driver might have and will ensure that the collected data are reflecting the normal driving behavior.

Despite all the advantages the NDS offer for studying the naturalistic driver behavior, other methods serve specific purposes. For instance, driving simulator studies could be chosen to test certain situations which might involve a high risk on the drivers. Also, such a situation could need specific technical and financial support that are not feasible in the NDS (Purucker et al., 2018). Same for on-road experiment and test track studies, the purpose of these studies will direct the decision towards which method to use. For example, the on-road experiment could be used as a proof of concept for certain technologies as an initial stage before the wide implementation of this technology in the automotive industry (Donmez et al., 2007). Also, test track studies could be used for the same purpose of designing/testing new technologies which relates to drivers' safety, such as FCW (Kiefer et al., 1999).

Table 1 summarizes and compares the descriptions of the above-discussed driver behavior measurement methods in terms of the study route, test duration, driving context including vehicle, driver, and infrastructure, and measures extraction.

Table 1 *A comparison between different driver behavior measurement methods*

	NDS	**On-Road Experiment**	**Test Tracks**	**Driving Simulators**
Study Route	Flexible, not specified, and covers the entire network	Specific route, which is identified based on the study purpose	Within the test track facility only	Hypothetical route with real scenes of the buildings and infrastructure
Test Duration	Long periods (i.e., months)	Day-long data collection, but the data could be collected on several days	Hours or less than an hour	Hours or less than an hour depends on the participant ability (i.e., simulator sickness)
Driving Context				
Vehicle	Driver's own vehicle	An instrumented vehicle which should be returned to the research center at the end of each day of the data collection	An instrumented vehicle used only within the test track facility	Ranges from desktop driving simulators to multi-axis motion simulators with 360 degrees of display
Driver	No instructions	The drivers are instructed to perform certain actions/driving style during the experiment	The drivers are instructed to perform certain actions/driving style which could include high risks, and they know that they are in a test and it could be restarted	The drivers are instructed to perform certain/driving styles or test new technologies, and they know that the test could be restarted

Table continues in the next page >>

	NDS	On-Road Experiment	Test Tracks	Driving Simulators
Infrastructure	No restriction on the type of the roadway facility (i.e., the drivers are free to use any roadway based on their own preferences and their destinations)	The drivers are instructed to drive on a specific route on a predefined roadway (e.g., freeway), and the predefined route could include one or more intersections	The drivers are limited to the test track facility type either the facility includes intersections or straight or curved segments	Based on the modeled scenario, the experiment could include a mix of intersections and segments or highway segments only
Samples	Large number of drivers could participate in a study (could reach more than 3000)	Limited number of drivers could participate depending on the number of the used instrumented vehicles	Limited number of drivers could participate depending on the number of the used instrumented vehicles	Limited number of drivers could participate
Measures Extraction	Long extraction, data processing, and incident identification process and high risk of data losses due to malfunctions of the equipment	Low data processing effort since the data collection could be triggered at certain maneuvers with a high risk of data losses	Low data processing effort since the test is short in time with a low risk of data losses	Low data processing efforts and low risk of data losses

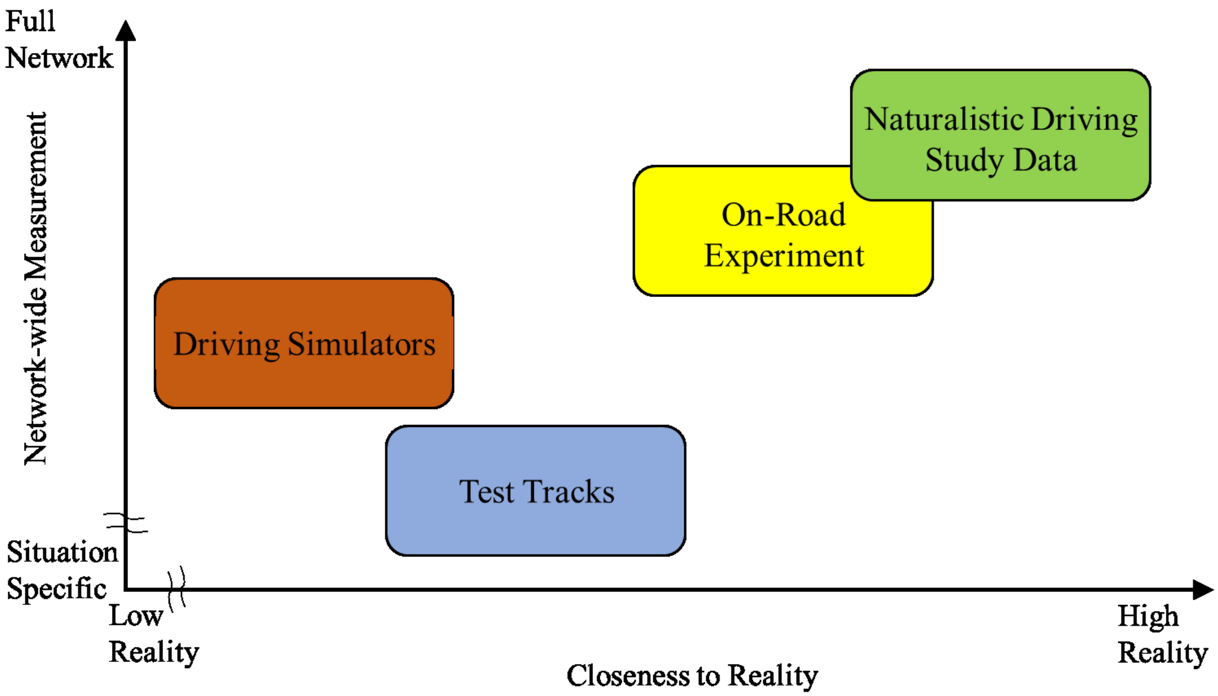


Figure 3. The relation between different behavior measurement methods according to network-wide measurement and closeness to reality.

Based on the previous discussion and the comparison in Table 1, the NDS data is the best source for the normal driving behavior. Moreover, the NDS data is not limited to a specific route so it could be used for network-wide behavior investigation. As shown in Figure 3, the NDS data are the most reliable and robust data in terms of closeness to reality where the drivers are using their own vehicles, are not committed to certain driving style, and are not subject to the test environment that the other measurement methods impose. Furthermore, the NDS data is spread over the network covering all the network features (e.g., intersections, segments, curves, ramps, etc.), which makes the NDS the best to provide network-wide behavioral measurements.

Therefore, Figure 3 places the NDS data at the high end of the spectrum for reality and network-wide measurement when compared to other behavior measurement methods. As presented in this figure, the driving simulator has the least degree of reality when compared to other measures. On the other hand, the test track studies are the most situation-specific method, and the driving simulators could provide more generic situations than the test tracks due to the modeling flexibility. Notice in Figure 3 that there is an overlap between the driving simulators and the on-road experiments in terms of the network-wide measurement. This overlap is because both measurement methods (i.e., the on-road experiment and the driving simulators) are limited to specific routes, either due to the modeling of the road environment in the simulators or due to the limits of the study route. Nevertheless, the on-road experiments in some cases could be extended beyond the simulators modeling capacity. Moreover, there is another overlap between the on-road experiments and the NDS since the driver in both methods is using a vehicle on the road with the real traffic. Nonetheless, the on-road experiments are limited to a specific route, and the driver is driving an instrumented vehicle in most cases, and sometimes an observer joins the driver during the experiment.

In conclusion, the NDS data cover the whole road network and is not limited to a specific route which means that the driver has the ability to drive through intersections and on segments without instructions. This makes the NDS data the best for the purpose of this dissertation, since it will allow the comparison of uninstructed driver behavior at the proximity of intersections and on segments. This advantage will also allow and facilitate reaching the other objectives of this dissertation.

3.2 NDS AND DRIVER BEHAVIOR STUDIES

NDS data has been used in multiple applications in driver behavior investigation and traffic safety. For instance, the "100-Car Naturalistic Driving Study", which is one of the most utilized naturalistic datasets, is considered one of the earliest studies to attempt the collection of large scale driving data (Dingus et al., 2006). This study had several goals, such as studying factors contributing to rear-end and lane-change collisions while characterizing drivers' inattention in relation to near-misses and collisions. The data was collected for 78 privately owned cars and 22 leased cars over a 12-to-14 month period for each vehicle. The Strategic Highway Research Program 2 (SHRP2) NDS, which is considered the largest NDS and the first large-scale study to investigate detailed estimates of crash risk, was aimed at preventing collisions or at least reducing their severity to minimize traffic injuries and fatalities (Dingus et al., 2015). The SHRP2 collected continuous driving data for almost 3 years for around 3,000 light vehicles drivers in six different locations around the US. In a similar NDS collection effort, the Canada Naturalistic Driving Study (CNDS) was conducted for more than 120 drivers in the City of Saskatoon, SK ("Canada Naturalistic Driving Study," 2018). In Europe, NDS data was collected for about 190 passenger cars drivers and 46 truck drivers. The main aims of this data collection effort were to analyse crash causation factors, driver distraction, and eco-driving (Bärgman et al., 2017).

In addition to the studies discussed in the literature review, several studies have utilized NDS data to investigate the driver behavior from various aspects. For example, using 100-Car NDS data, Chen et al. (2015) investigated driver behavior during lane-changing (i.e., overtaking maneuvers) events. This study showed that the frequency of lane-changing differs with various speed groupings and the minimum TTC increases with a rise in travel speed. Wu and Thor (2015) suggested an approach to compare crash sequences and to quantify crash risk. Unfortunately, only a small sample size - two rear-end events - was used in the analysis. Carney et al. (2016) investigated 400 rear-end collisions involving teen drivers. The following performance measures were investigated: driver distractions (i.e., attending to a passenger, cellphone use, attending to a moving object, etc.), eye-off-road time, and response times to braking lead vehicles. The results showed that cell phone usage was the most frequent cause of rear-end collisions. Moreover, half of the teen drivers did not brake properly or steer before the impact while using cell phones. Lin et al. (2008) analyzed the causes of rear-end conflicts using naturalistic data collected for 50 taxi drivers over a 10-month period in Beijing. It was concluded that tailgating was the most common contributing factor to rear-end conflicts.

Ye et al. (2017) detected drivers' distraction due to the engagement in secondary tasks while driving. Five behavioral measures were extracted from SHRP2 NDS, namely, speed, longitudinal and lateral acceleration, yaw rate, and throttle position. These behavioral measures were used to train three artificial neural networks to detect secondary tasks such as calling, texting, and interaction with passengers. The results indicated that these measures were capable of detecting these tasks with high accuracy (Ye et al., 2017). Moreover, using the same NDS, the impact of the adverse weather conditions on the driver behavior was investigated (Ahmed and Ghasemzadeh, 2018). The driver behavior was described in terms of headway and speed only. The results showed that a significant difference between driver behavior in clear and rainy weather. Furthermore, the drivers were more likely to drive below the speed limits on road segments of high posted speed limits (Ahmed and Ghasemzadeh, 2018).

Using a dataset of SHRP2 NDS, Arbabzadeh and Jafari (2018) classified the driving outcomes in terms of normal driving, near-crashes, and crashes. The normal driving was split into three types based on executing unsafe driving behavior maneuvers and engagement in a secondary task. Several variables were used to train and validate a multinomial logistic regression model including drivers' age, gender, driving experience, cell phone use, impairment, seat belt usage, passenger in the adjacent seat, speeding, weather, road surface condition, lighting, roadway alignment (i.e., curve or straight segment) and grade, roadway facility type (e.g., two-way two-lane roadways, divided and undivided, etc.), traffic flow condition (e.g., unstable flow, free flow, restricted flow with leading vehicle, etc.), traffic control (i.e., school zone, construction zone, traffic signal, no traffic control, etc.), and land use (e.g., residential area, school, business/industrial, etc.). It is worth mentioning that all the variables used were either binary or categorical variables. The identification accuracy of the model 61% and 75.6% for near-crashes and crashes. On the other hand, the model had a significantly high false alarm rate for normal driving behavior identification (Arbabzadeh and Jafari, 2018).

Again, using a subset of SHRP2 NDS, the driver behavior in terms of lane keeping behavior was explored in heavy rain condition (Ghasemzadeh and Ahmed, 2017). A limited number of matched cases of driver behavior in clear and heavy rain were compared. The matched cases were identified in this research as the cases when the same driver uses the same route in clear and rainy weather. The results revealed that the heavy rain could increase the standard deviation of lane position. Furthermore, the drivers showed an increase of their lane keeping behavior in heavy rain while driving on a high speed roadways (Ghasemzadeh and Ahmed, 2017).

Wege et al. (2013) assessed truck drivers' response to an FCW in terms of brake reactions and visual attention allocation. The data was first collected for several days without sending a warning to the driver but recording it to a file. The same data was then collected while the warning was activated and sent to the driver when necessary. The main finding of this study was that the FCW increased drivers' visual attention allocation towards the road during the rear-end event. However, after the threat ended, drivers' visual allocations turned to the warning source.

McLaughlin et al. (2009) developed an evaluation methodology, which measures the performance of FCW algorithms using 100-Car NDS. Seventy collision scenarios, which were focused on rear-end collision and near-collision avoidance with a lane-change, were inputted into three independent algorithms. The alert timing and driver reaction time were collected to analyze the algorithms' performance for three different levels of braking (0.5 g, 0.675 g, 0.85 g). Results showed that all three algorithms provided alerts more frequently than was acceptable by drivers.

Alden et al. (2016) proposed a methodology to collect naturalistic driving data for drivers in locations with a high occurrence of animal-related collisions in order to investigate driver and animal behavior characteristics before and during animal-vehicle encounters. For motorcycle-related collisions, Williams et al. (2015) conducted an explanatory analysis using 100 motorcyclists riders NDS. The data was extracted from videos by orienting and modifying a previously developed video reduction dictionary to identify motorcycle events. The preliminary analysis identified 22 collisions, most of which were low-speed ground impacts.

Lee and Yeo (2016) suggested a rear-end collision warning system based on drivers' characteristics. These characteristics were defined as the following vehicle's speed and acceleration, the leading vehicle's acceleration, the gap distance and the relative speed between the leading and the following vehicles. To develop the proposed algorithm, the Next Generation Simulation (NGSIM) trajectory data was used as a source for drivers' characteristics. It is worth mentioning that the proposed algorithm was evaluated using only 53 events of interest out of 322 car-following cases identified in the NGSIM dataset.

Wang et al. (2013) developed a driver model to simulate the throttle and braking of the driver. As an application of this model, an FCW and Avoidance algorithm was proposed. To develop this model, driving data was collected during an earlier study for 33 nonprofessional drivers using a data collection system installed in a test vehicle. The data was collected in a real traffic environment; however, the test duration was around two hours only (Chi et al., 2010). The algorithm, proposed by Wang et al. (2014), was designed to relay four levels of warning combined with automatic braking activation. The warning and braking levels ranged from no warning and no braking to emergency warning and automatic braking activation when the driver is not braking. The thresholds for the warning levels were defined based on a TTC range, and the braking activation was based on the desired brake pressure.

Young et al. (2007) designed and tested several warning applications, including an FCW, to assess the impact of such applications on driver car-following behavior, mental workload, and system acceptance. The FCW was designed to trigger several warning levels based on the time gap between the host and the leading vehicles. To assess the impact of the FCW system, time gap data was collected from a road test, which involved 23 drivers (only 17 drivers completed the designed test). To evaluate the impact of the FCW on the drivers' following behavior, the average, minimum, and standard deviation of the time gap were compared with and without the FCW system. In general, with the FCW system, the drivers had a larger time gap than without the system (Young et al., 2007).

The NDS data was used to characterize drivers' overtaking maneuvers (Chen et al., 2015), distraction detection (Ye et al., 2017), lane keeping behavior (Ghasemzadeh and Ahmed, 2017), and drivers outcomes classification (Arbabzadeh and Jafari, 2018). Moreover, the NDS data was used in traffic safety to investigate the contributing factors in crashes for different drivers, including young drivers (Carney et al., 2016; Lin et al., 2008; Wu and Thor, 2015). The NDS data has not only contributed to the development of DAS, such as FCW systems (Lee and Yeo, 2016; Tawfeek and El-Basyouny, 2018b; J. Wang et al., 2013), but also to testing these technologies (McLaughlin et al., 2009; Tawfeek and El-Basyouny, 2018b; Wege et al., 2013). In summary, the NDS data has provided valuable insights into different aspects of drivers' behavior.

3.3 NDS DATA DESCRIPTION

NDS was carried out to investigate driver behavior and performance by collecting detailed data on volunteer drivers, vehicles, and the surrounding environment during normal every day driving. The data collection equipment was typically hidden, and no specific instructions were given to the volunteer drivers (Mclaughlin et al., 2009; UDRIVE, n.d.). The NDS was collected as part of the Safety Pilot Model Deployment (SPMD). SPMD was a data collection effort under real-life conditions with about 3,000 vehicles equipped with V2V communication devices in Ann Arbor, Michigan, US (Henclewood and Rajiwade, 2015). SPMD was part of the "Connected Vehicle Safety Pilot Program" research initiative which was aimed at evaluating the safety benefits of CV technologies. The dataset comes from the Data Acquisition System, which was developed by the Virginia Tech Transportation Institute. This dataset is available on the ITS Public Data Hub, which is operated by the Federal Highway Administration (FHWA) under the US DOT (USDOT, 2018). The dataset was collected from 64 host vehicles equipped with Integrated Safety Devices (ISD) for two months (October 2012 and April 2013). For each second, the ISD unit recorded data including position and GPS details (e.g., latitude, longitude, elevation, speed, and GPS-based data fidelity measures) and driver behavior data (e.g., braking status, speed, yaw rate, and acceleration). In addition, the dataset contains radar data collected from the ISD unit. The radar data provided information pertaining to the detected vehicle type (e.g., light vehicle, heavy vehicle, bike, pedestrian, or unclassified), status (i.e. moving and in path), the distance from the host vehicle, and the relative speed to the object in front of the host vehicle (Henclewood and Rajiwade, 2015). The data was stored in two *.csv* files with a total file size of around 40.7 GB. The first file housed the information related to the vehicle operation and position for the host vehicle. The second file stored the radar data of the detected vehicles around the host vehicle. After fusing these two files together, more than 150 million records were analyzed as will be explained later.

3.4 Identifying of Car-Following during Braking Events

In order to quantify and make a comparison between driver behaviors at different road network locations, specific events should be extracted from the NDS to represent the driver behavior at various locations (e.g., at intersections and on roadway segments). These specific events were identified as the car-following events during braking since the car-following is the most common interaction on roadways and braking maneuvers are the most frequent driver response to critical safety incidents in car-following situations as discussed in Chapter Two. Moreover, behavioral measures at both locations (i.e., at intersections and on roadway segments) should be extracted or estimated.

Despite the usefulness of the NDS data, it can also be quite challenging since a significant amount of data manipulation is required. This is a common issue with big NDS data especially in the absence of video footage. Consequently, a car following identification algorithm could be used to select targeted events. Kusano et al. (2014) developed such an algorithm to identify host vehicle braking events while following another vehicle. The algorithm successfully identified approximately 92% of all car following events when compared to manual inspection of video log data in a100-Car NDS. The same algorithm was used to characterize driver lane changing behavior by studying the minimum TTC at various speeds and to compare TTC and Enhanced TTC thresholds for triggering Forward Collision Warning (FCW) (Chen et al., 2016, 2015). Moreover, the algorithm was utilized to compare the TTC for different speed bins or different drivers' demographic groups (i.e., age and gender) in car-following scenarios while braking (Kusano et al., 2015; Montgomery et al., 2014). Also, Tawfeek and El-Basyouny (2018b) and Tawfeek and El-

Basyouny (2018c) adapted this algorithm for the identification of the car-following events in NDS data to develop a perceptual FCW. Consequently, the algorithm developed by Kusano et al. (2014) will be adapted to process the NDS and to identify incidents of interest to serve the objectives of this dissertation. The following steps, which were coded in MATLAB, summarize the NDS data processing and events identification. Moreover, the entire data processing framework is summarized in Figure 5.

1. Preparing and filtering the data
    a. Remove all sensed object types except those which were recognized as light vehicles (i.e., passenger cars). For instance, the objects that were detected as bikes, pedestrians, and heavy vehicles (i.e., trucks) were removed. This was done to limit the scope of the study to focus entirely on passenger cars interactions since a car-following-truck behavior is different from that of a car-following-car behavior (Aghabayk et al., 2012; Sarvi, 2013).
2. Identifying car following events during braking using Kusano et al. algorithm. For further details, readers are referred to (Kusano et al., 2014)
    a. Identify all braking events for the host vehicle only when the vehicle's speed is above 10 mph (4.47 m/s) for at least 1 second of the braking time.
    b. Identify objects in front of the host vehicle, during the braking events, with headway less than 3 seconds for at least 15% of the braking time and within 2 m laterally from the host vehicle's path for at least 65% of the braking time.
    c. Remove fixed objects, parked vehicles, and slow vehicles in adjacent lanes in two steps. First, candidate leading vehicles should have (i) speed higher than 3 mph (4.8 km/h) and moving away from the host vehicle, or (ii) speed less than 3 mph

and moving away or closer to the host vehicle, and the minimum distance between the vehicles is 6 m, or the host vehicle has a speed less than 30 mph (48.28 km/h). Second, candidate leading vehicles should also have a change in azimuth angle less than 5 degrees/second. When the change in azimuth is between 3 and 5 degrees/second, the values within this range should be only 20% of the braking time and 35% of the braking time if the braking time was less than 1 second. It is worth mentioning that the azimuth is defined as the angle between the host vehicle path and the candidate leading vehicle.

   d. Select the leading vehicle if it is, on average, within 1.5 m laterally from the host vehicle path and was the closest to the host vehicle.

3. Selecting incidents of interest
   a. Scan each identified following event for the minimum following distance.
   b. Extract and calculate driver behavior measures occurring at the minimum following distance while braking, namely, TTC, time headway, vehicle speed, relative speed, vehicle acceleration, and jerk.

where TTC is calculated as the ratio between the range (i.e., following distance) and the relative speed. The TTC is defined as the time left for two vehicles to collide if no evasive maneuver is performed by at least one of them (Hayward, 1972; Hydén, 1987). The TTC could be estimated based on the following equation.

$$TTC(\text{seconds}) = \frac{r}{v_F - v_L} \qquad (1)$$

where r is the following distance in meters, $(v_F - v_L)$ is the relative speed, v_F is the speed of the following vehicle, and v_L is the speed of the leading vehicle as shown in Figure 4.

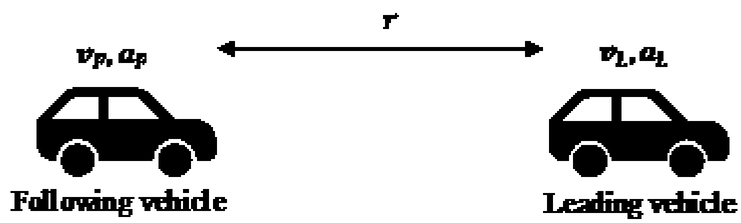


Figure 4. The main parameters of a car-following situation.

The longitudinal jerk, which will be referred to as jerk, is the change of the acceleration over the time and could be represented in the following equation:

$$Jerk\ (m/s^2/s) = \frac{da_F}{dt} \qquad (2)$$

where da_F is the difference between the following vehicle acceleration in two consecutive time steps and dt is the time step length. It is worth mentioning that the lateral jerk was not included in the analysis since the used dataset lacks the information needed to calculate it.

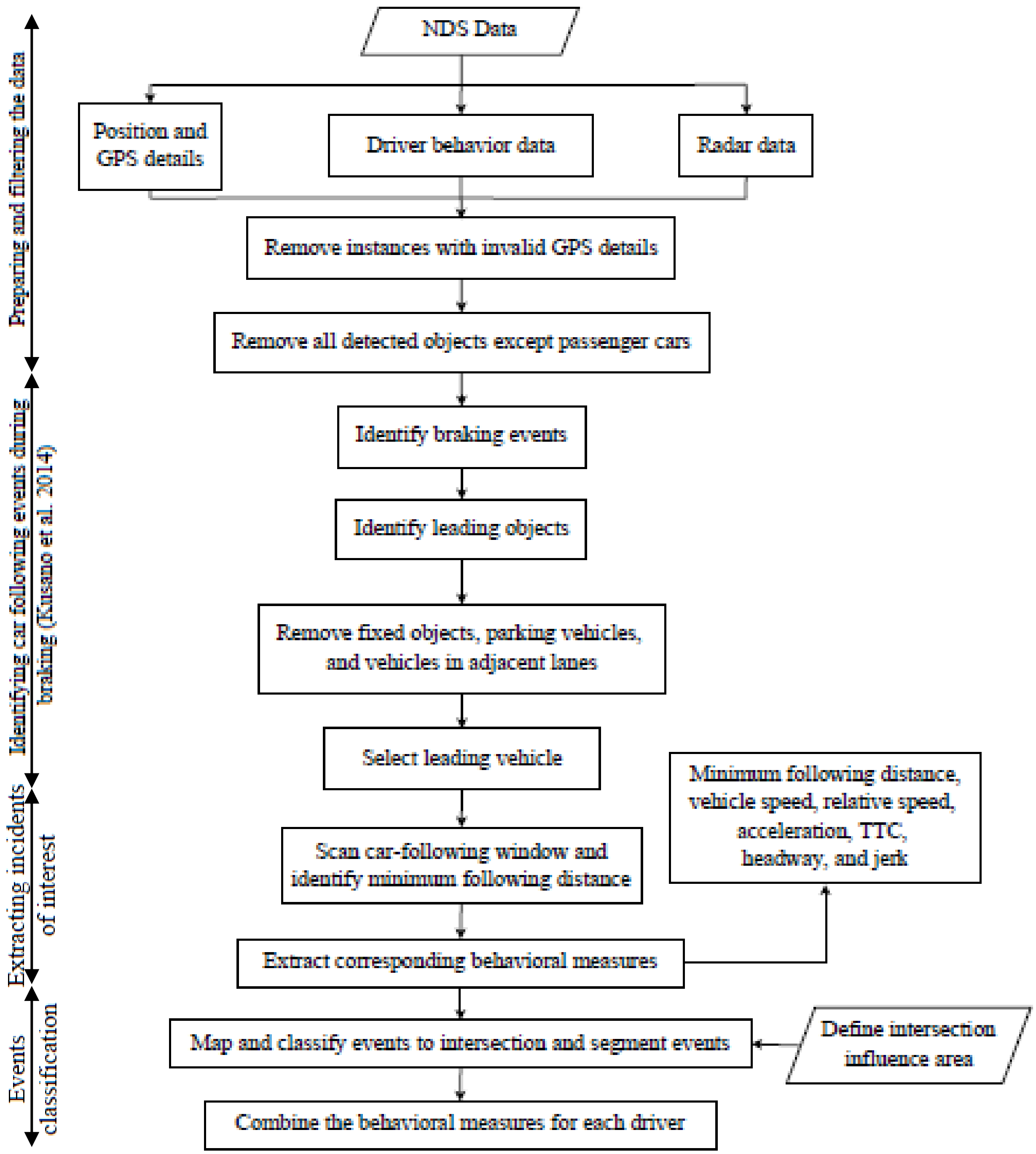


Figure 5. Events of interest and behavioral measures extraction framework.

3.5 INTERSECTION INFLUENCE AREA DEFINITION AND EVENTS MAPPING

As shown in Figure 5, after extracting the events of interest, ArcGIS was used as a tool to map the identified car following events during braking and to classify each event into either an intersection- or segment- related event. The events of interest were imported to ArcGIS using the available position data of the host vehicle (i.e., latitude and longitude). Before starting the classification process, intersection- and segment- (non-intersection) related events should be defined. The current literature has various definitions of the intersections influence area. For instance, both the American Association of State Highway and Transportation Officials (AASHTO) and Highway Safety Manual (HSM) defines at-grade intersections by physical and functional areas (AASHTO, 2011, 2010). The physical area is the area bounded by the stop lines of both intersecting roads. On the other hand, the functional area extends upstream the intersections on the intersecting roads. This area includes the driver's decision and maneuver distances and queue storage distance. More specifically, intersection-related crashes are defined as the crashes which occur within 15 to 152 meters from the intersection center point (FHWA, 2009; R.Stollof, 2008). On the other hand, the 152 meters intersection influence area was considered to be suitable only for intersections approaches according to the 40 mph speed limit (Cottrell and Mu, 2005). It is worth mentioning that intersection crashes and intersection-related crashes have been defined as the crashes that occurred upstream or inside the physical area of the intersection (Wang et al., 2009). Therefore, similar to the definition of intersection crashes, the events of interest were identified as intersection events when they occurred either upstream or inside the physical area of the intersection.

Since speed is the main contributor to a driver's decision (i.e., perception-reaction) distance and maneuver (i.e., braking) distance, the speed limit should be considered when deciding the influence area (Fitzpatrick et al., 2000). Consequently, the events of interest were split into three categories based on the host vehicle speed, namely; low, medium, and high speeds. The boundaries of low, medium, and high speeds were defined as less than 25 mph (40 km/h), between 25 mph and 55 mph (88 km/h), and more than 55 mph, respectively. These boundaries are chosen based on the speed limits in the City of Ann Arbor, Michigan (City of Ann Arbor, 2013). Based on these categories, intersections influence area for each category was set and defined as shown in Figure 6.

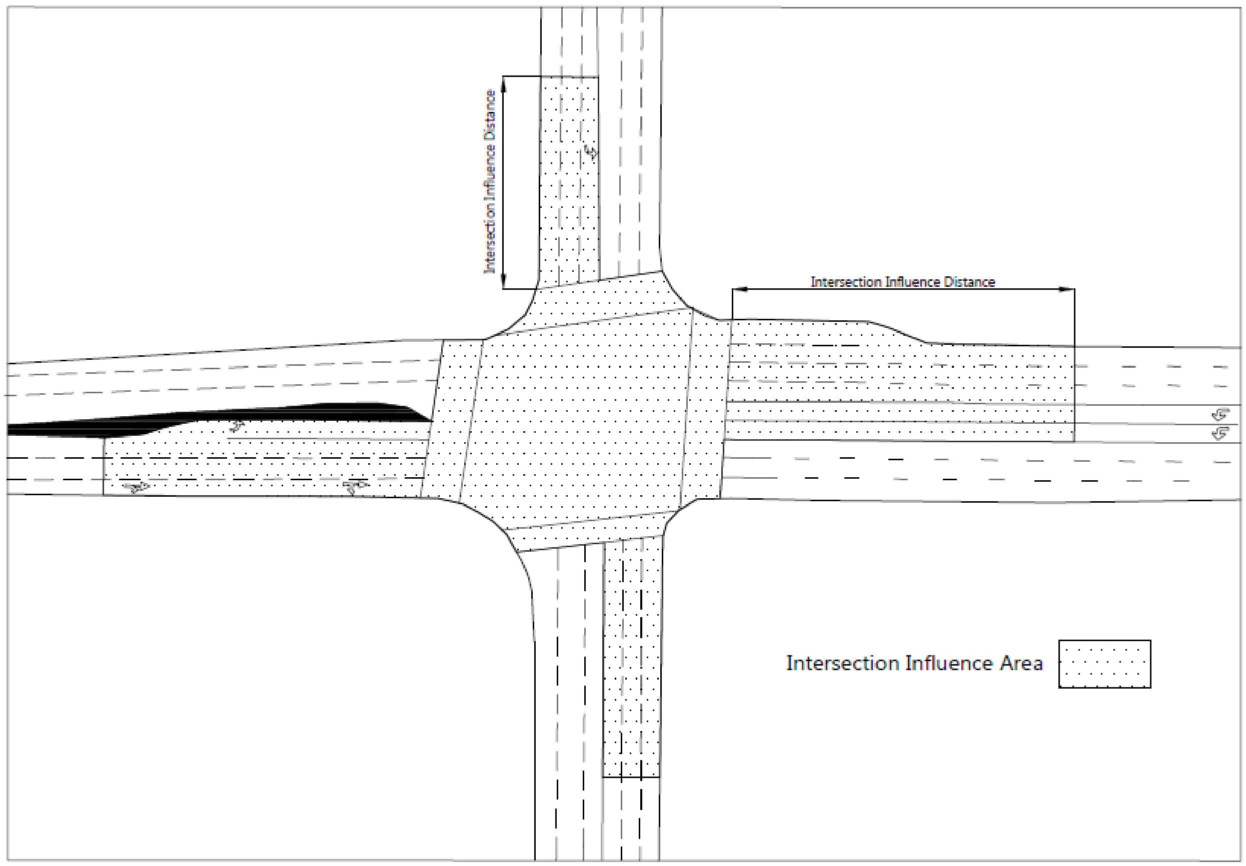


Figure 6. Sketch of intersection influence area.

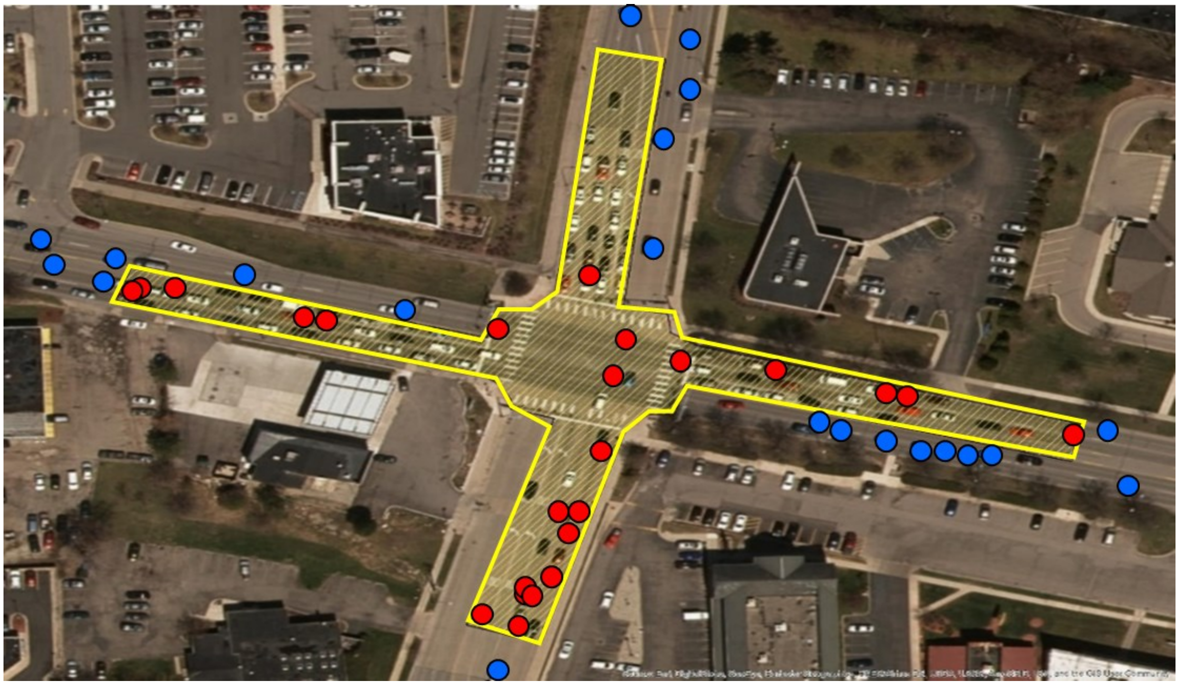


Figure 7. Example of the intersection events (red dots) and segment events (blue dots).

The intersection influence distance is calculated based on the driver's Stopping Sight Distance (SSD), which include driver's decision and maneuver distances as shown in the following equation.

$$SSD\ (meter) = 0.278 \times v \times PRT + \frac{v^2}{254\left(\frac{a}{g} \pm G\right)} \qquad (3)$$

where v is vehicle speed in km/h, PRT is driver's Perception-Reaction Time in seconds, a is vehicle deceleration in m^2/s, and G is roadway longitudinal slope.

The SSD was estimated for low, medium, and high-speed categories as 30, 80, and 150 meters respectively based on AASHTO recommendations for PRT of 2.5 seconds and deceleration of 3.41 m^2/s (AASHTO, 2011). Therefore, the intersection influence distance shown in Figure 6 will be 30, 80, and 150 meters for low, medium, and high-speed categories. Based on this distance, the intersection influence area was determined, and any event of interest located within this area was classified as an intersection-related event or, otherwise, considered as a segment-related event as shown in Figure 7. A total of 44,383 events were extracted, based on the procedure presented in Figure 5, with an average of 704 events for each driver. These events were classified based on the position of the host vehicle as mentioned before and the breakdown of the events based on the position is shown in Table 2. From this table, the percentages of the intersection and segment events were 30.2% and 69.8%, respectively.

Table 2 *Events breakdown by event type and speed group*

Speed Group	Intersection Events	Segment Events	Total per Speed Group
Low	11403	14962	26365
Medium	1995	12614	14609
High	0	3409	3409
Total Events	13398	30985	44383

Figure 8, Figure 9, and Figure 10 all show the city of Ann Arbor map after classifying the events into intersection- and segment- related events for the three above-mentioned speed groups. As shown from the figures and from Table 2, the high-speed group does not include any intersection events because the lower limit for this speed group is 88 km/h (55 mph) and it is difficult to have a braking event at an intersection with such a high speed.

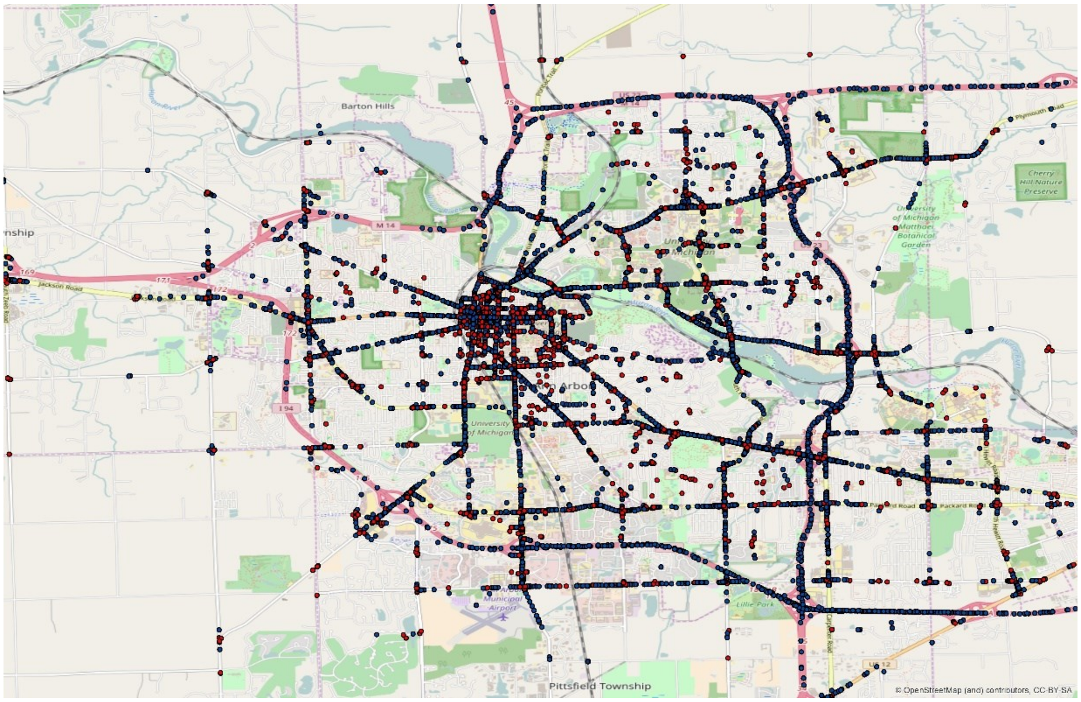


Figure 8. The distribution of intersection (11403 events) (red dots) and segment (14962 events) (blue dots) car following events while braking in the low-speed group.

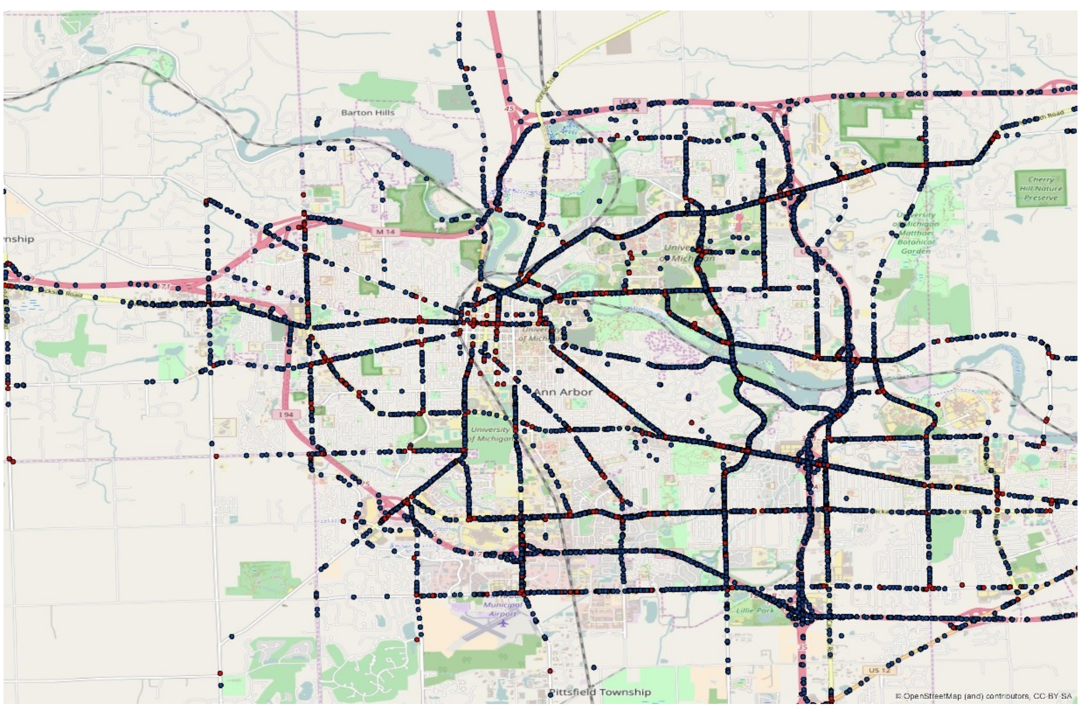


Figure 9. The distribution of intersection (1995 events) (red dots) and segment (12614 events) (blue dots) car following events while braking in the medium speed group.

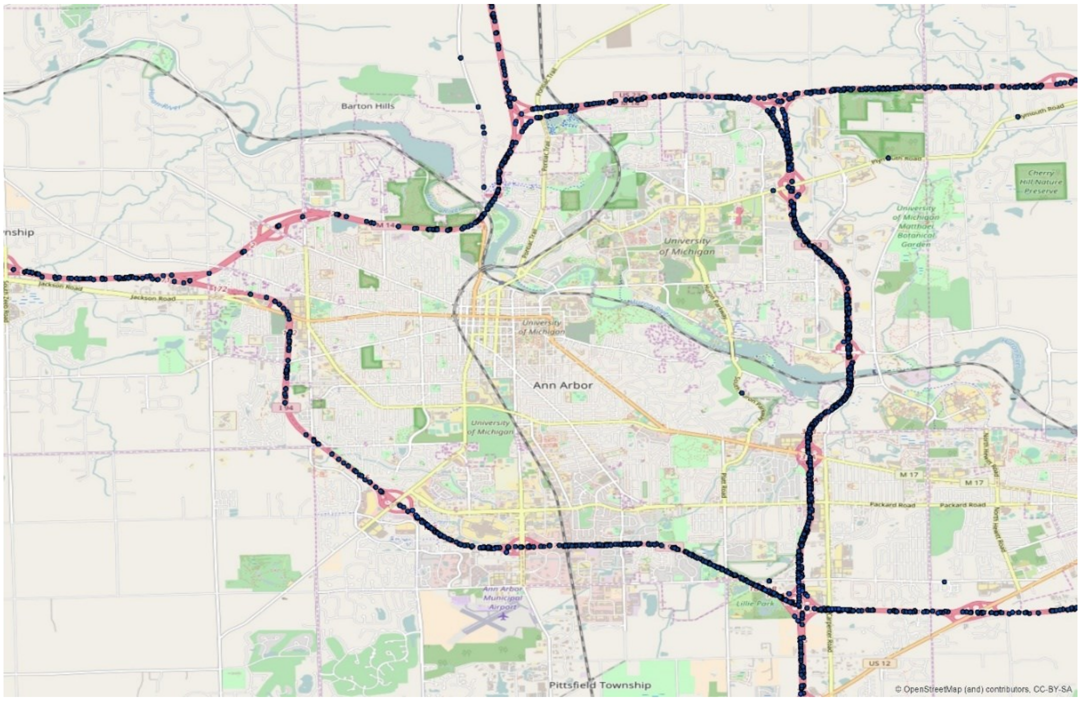


Figure 10. The distribution of segment car following events (3409 events) (blue dots) while braking in the high-speed group.

Moreover, Figure 8, Figure 9, and Figure 10 showed the macroscopic spatial distribution of the events with respect to the speeds of the events. As shown in the figures, the events of the low-speed category were more centered around the downtown core. However, there were some events scattered on the highways within the city boundary and on the ring road. This reflected that those scattered events took place in a more congested situation when compared to the other speed categories. On the other hand, for the medium speed category showed in Figure 9, the events were more clustered on the arterials and main highways. This distribution was expected because of the speed range of this category (i.e., medium-speed category). Finally, for Figure 10 and the high-speed category, the events were only on the ring road and the surrounding rural highways. This spatial distribution highlighted that the speed categories not only supported the definition of the intersection influence area but also separated the driver behavior based on the traffic flow condition. It is worth mentioning that the speed was used as a traffic flow indicator for context-aware DAS (Bejani and Ghatee, 2018).

3.6 BEHAVIORAL MEASURES QUANTIFICATION

After the classification of each event (i.e., segment or intersection) in ArcGIS, the results were exported from ArcGIS to be processed in MATLAB (The Mathworks Inc., 2016). The probability density functions for acceleration, following distance, TTC, relative speed, and jerk were developed in each speed group for both intersection- and segment-related events. The actual values of the above-mentioned measures were fitted to 17 continuous distribution types, namely; Beta, Birnbaum-Saunders, Exponential, Extreme value, Gamma, Generalized Extreme Value, Generalized Pareto, Inverse Gaussian, Logistic, Log-logistic, Log-normal, Nakagami, Normal, Rayleigh, Rician, *t* Location-Scale, and Weibull.

Table **3** shows the distribution's name and parameters, which best fits the actual values of the behavioral measures (i.e., acceleration, minimum following distance, TTC, relative speed, and jerk) for each event type (i.e., intersection- and segment- related events) and each speed group (i.e., low, medium, and high speeds). The distribution was considered to best fit the actual values of the behavioral measures since the Bayesian Information Criterion (BIC), and the Akaike Information Criterion (AIC) had the lowest values. Moreover, Figure 11, Figure 12, Figure 13, Figure 14, and Figure 15 show the plot of the probability density functions, which best fits the actual values for all the cases.

Since the distribution of all the measures was not normally allocated, a nonparametric statistical test (i.e., Kolmogorov-Smirnov (KS) test) was used to check the significance of the difference between the intersection and segment event distributions. The null hypothesis is that the values for each measure of the intersection and segment-related events are from the same continues distribution. As shown in

Table **3**, the results of KS test revealed that all the measures of intersection events were different from segment events at 95% confidence level except the jerk in medium speed group. It is worth mentioning that the *p*-value was estimated at 0.0514, which is significant at 90% confidence level, when comparing the jerk for intersection and segment events in this speed group. These results indicate that driver behavior at intersections is substantially different from driver behavior on segments. This difference was also evident in Figure 11, Figure 12, Figure 13, Figure 14, and Figure 15. Such results are crucial when developing any context-aware DAS because drivers will receive the warning differently in varied contexts. In other words, drivers could accept a warning relayed to them on midblock segments while considering the same warning at intersections as an early warning which will have an impact on the entire system acceptance.

Table 3 also shows the mean values of the selected behavioral measures in the third column between parentheses. For the jerk, the mean was calculated for the negative and positive jerk values to provide a complete picture about this behavioral measure since the average of all values will be around zero, as shown in Figure 15, and will not give any useful information. However, the distributions' parameters of the jerk (in the fifth, sixth, and seventh columns) were estimated based on all jerk values (i.e., negative and positive values). It is noteworthy that the negative jerk is more efficient in detecting the aggressiveness of the drivers when compared to the positive jerk (Feng et al., 2017). As shown in

Table **3**, measures' mean values for intersection-related events were less than segment-related events. For instance, the mean value of the deceleration of low-speed intersection events (i.e., 1.68 m/s^2) was higher than the deceleration of low-speed segment events (i.e., 1.43 m/s^2). This same trend was noticed as well for the values of following distance, TTC, relative speed, jerk, where all these measures of car-following during braking events at intersections were less than segment events. Although, the differences between the jerk values at intersections and on segments were not huge, they were statistically significant as discussed earlier. This indicated that the drivers tend to have fewer deceleration rates, larger following distance, less relative speed magnitudes, and higher negative jerk during braking when driving on segments than when driving at the proximity of intersections. Moreover, the mean values of the behavioral measures in the medium-speed group were larger than the values of the low-speed group.

In summary, drivers' following behavior during braking at intersections is statistically different from their behavior on midblock segments. In addition, drivers are more likely to follow closer to leading vehicles, to have higher deceleration rates, higher relative speed, lower TTC, and larger jerk at intersections. This means that the drivers are relatively more aggressive when approaching intersections than when driving on segments.

Table 3 *The parameters of the fitted distributions*

	Speed Group	Location (mean value)	Distribution Name	Parameters			*P-value*
				Location	Scale	Shape	
Acceleration (m/s^2)	Low	Int.(-1.68)	*t* Location-Scale	-1.667	0.745	20.687	<0.001
	Low	Seg.(-1.43)	Extreme Value	-1.060	0.663	-	
	Med.	Int.(-1.19)	Extreme Value	-0.847	0.608	-	<0.001
	Med.	Seg.(-1.00)	Extreme Value	-0.715	0.507	-	
	High	Seg.(-0.85)	Extreme Value	-0.624	0.405	-	-
Following Distance (m)	Low	Int.(9.01)	Generalized Extreme Value	7.127	3.153	0.020	<0.001
	Low	Seg.(11.05)	Generalized Extreme Value	8.496	4.071	0.047	
	Med.	Int.(19.5)	Gamma	-	4.430	4.403	<0.001
	Med.	Seg.(20.99)	Gamma	-	4.984	4.212	
	High	Seg.(29.07)	Birnbaum-Saunders	-	25.100	0.563	-
TTC (sec)	Low	Int.(10.48)	Generalized Extreme Value	3.556	3.081	0.739	<0.001
	Low	Seg.(17.23)	Birnbaum-Saunders	-	9.383	1.295	
	Med.	Int.(21.42)	Log-normal	2.549	1.087	-	<0.001
	Med.	Seg.(30.00)	Gamma	-	20.834	1.440	
	High	Seg.(31.81)	Birnbaum-Saunders	-	24.162	0.794	-
Relative Speed (m/s)	Low	Int.(-2.11)	Extreme Value	-1.368	1.212	-	<0.001
	Low	Seg.(-1.54)	Extreme Value	-0.919	0.964	-	
	Med.	Int.(-2.13)	Generalized Extreme Value	-2.208	2.340	-0.933	<0.001
	Med.	Seg.(-1.28)	*t* Location-Scale	-0.487	0.435	1.1189	
	High	Seg.(-1.23)	Generalized Extreme Value	-1.312	1.315	-0.870	-
Jerk ($m/s^2/s$)	Low	Int.(-1.47/1.88)*	*t* Location-Scale	0.059	1.082	2.145	<0.001
	Low	Seg.(-1.43/1.98)*	*t* Location-Scale	0.136	1.049	1.954	
	Med	Int.(-1.84/2.22)*	*t* Location-Scale	0.147	1.453	2.371	0.0514
	Med	Seg.(-1.79/2.14)*	*t* Location-Scale	0.169	1.299	2.139	
	High	Seg.(-1.78/2.19)*	*t* Location-Scale	0.266	1.529	2.787	-

*The average is calculated for (negative/positive) jerk separated

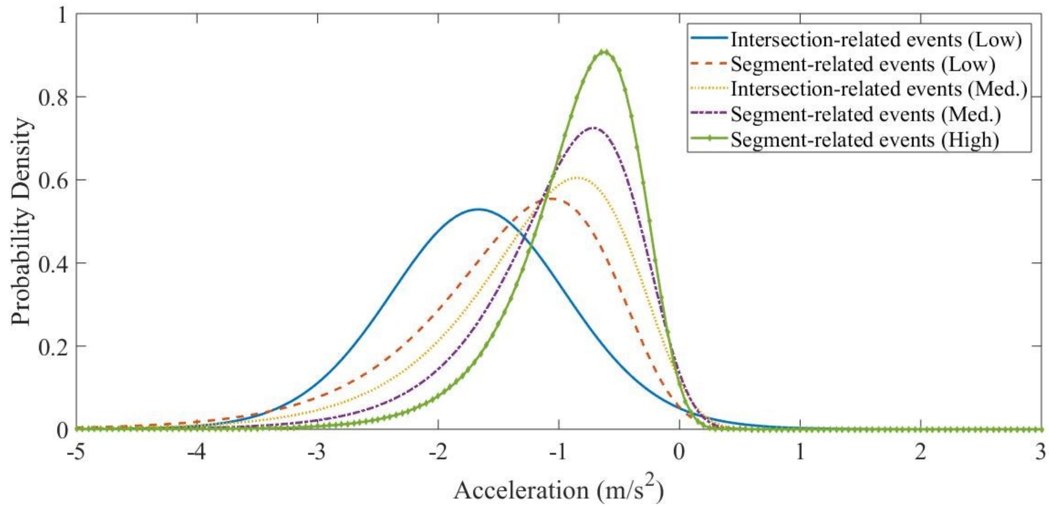


Figure 11. Probability Density Functions for the acceleration at different locations (speed category).

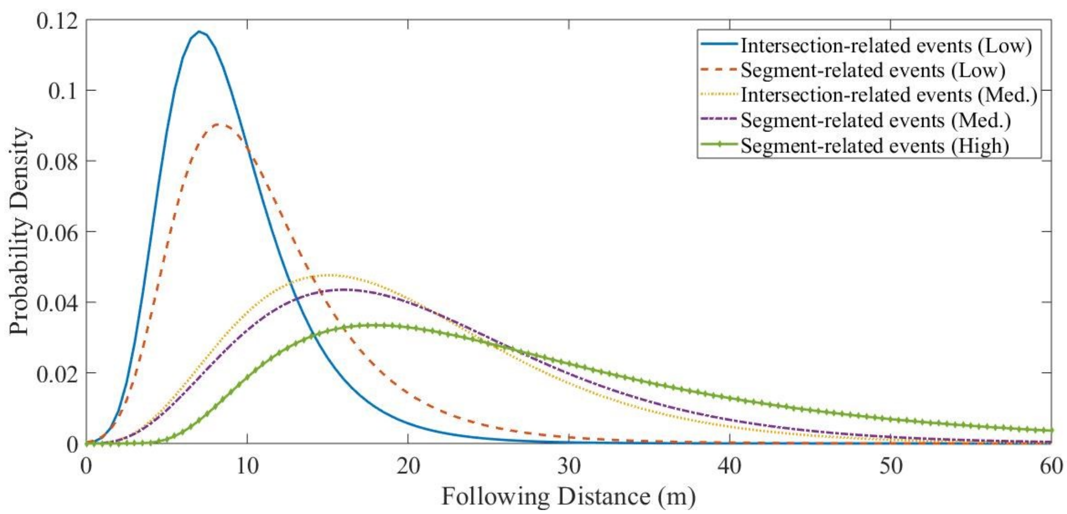


Figure 12. Probability Density Functions for the following distance at different locations (speed category).

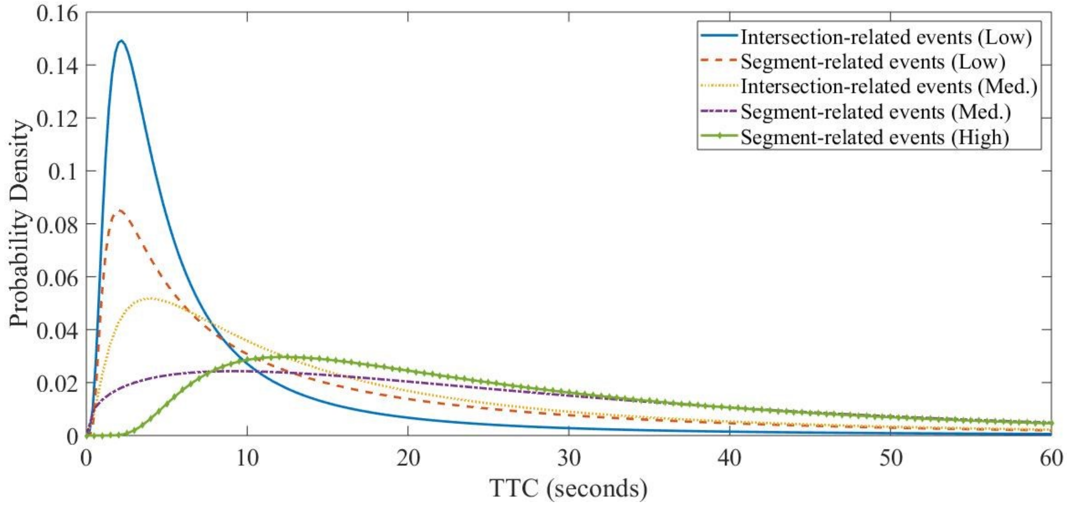


Figure 13. Probability Density Functions for the TTC at different locations (speed category).

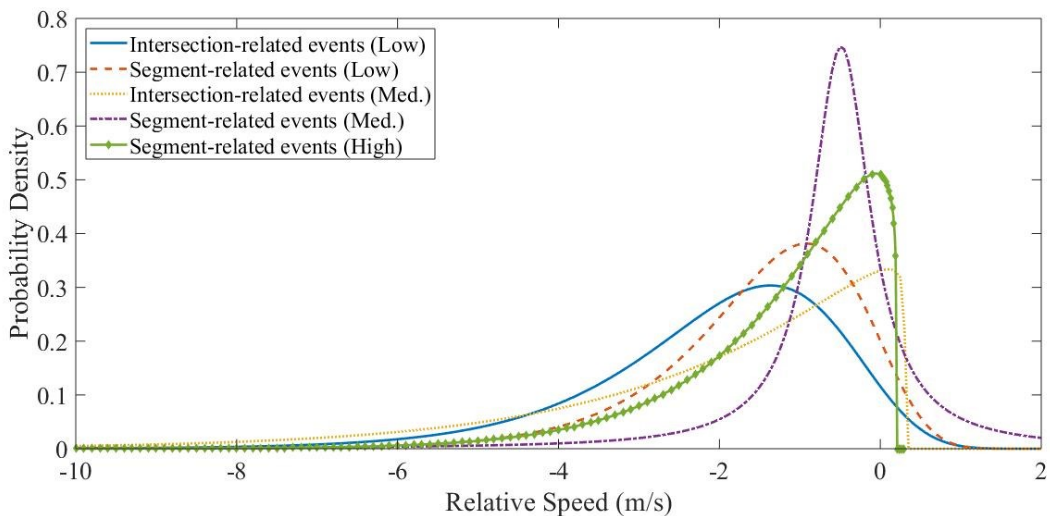


Figure 14. Probability Density Functions for the relative speed at different locations (speed category).

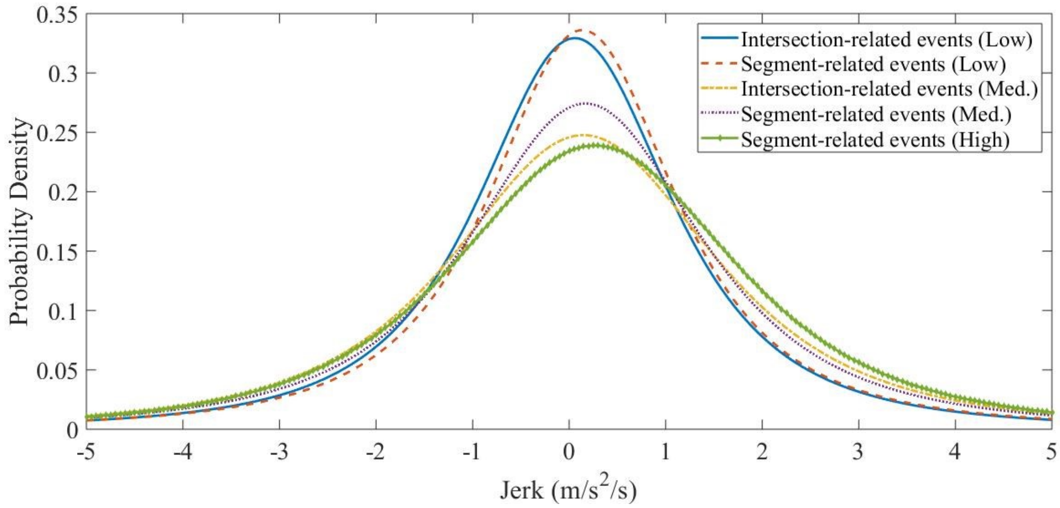


Figure 15. Probability Density Functions for the jerk at different locations (speed category).

3.7 Summary and Discussion

This chapter quantified five driver behavioral measures (i.e., acceleration, minimum following distance, TTC, relative speed, and jerk) for car-following during braking situations. This quantification is carried out for events occurring at intersections and midblock segments. The driver behavioral measures were extracted from the NDS.

The events of interests were extracted and onto the road network using ArcGIS to identify intersection and segment-related events. The intersection-related events were identified as the events that occur within the intersection influence area. This chapter defined the intersection influence area as the space where the driver behavior is affected substantially by an intersection ahead, which will impact the driver's maneuvering decisions including car-following, braking, and lane changing decisions. This influence area includes the physical area of the intersection, which is bounded by the stop lines and the area upstream the stop line of the intersection limited by the SSD, which is estimated based on the speed of the driver. This area was defined for three speed groups (i.e., low, medium, and high). Then, the probability distributions for the behavioral measures for each speed group for intersections and segments were developed and fitted to continuous probability distributions. In addition, intersection and segment driver car-following behavior during braking was compared by checking the significance of the difference between the behavioral measures distributions in each speed group.

The results revealed that there was a considerable difference between driver behavior during braking at intersections and on segments. This difference was defined as the differences between the behavioral measures distributions of intersection- and segment-related events. Using KS test, these differences were statistically significant at 95% confidence level. Moreover, this chapter documented the distributions' names and parameters that best fitted the behavioral measures. By investigating the mean values of the behavioral measures, it was found that drivers approaching intersections had higher deceleration rates, smaller minimum following distance, smaller TTC, higher relative speeds and larger jerk than when driving on segments. This insight implies that the drivers are more likely to exhibit aggressive behavior when approaching intersections. Considering the insights of this chapter, the proposed procedures in the literature could misclassify a driver as aggressive when most of the data was collected in high intersection density areas (e.g., urban areas or downtown areas). Therefore, it is crucial to consider the driver location when classifying the drivers based on their aggressiveness and incorporate this classification in the context-aware DAS which will be discussed later in this dissertation.

4. AN OVERVIEW ON THE CONTEXT-AWARE SYSTEM AND THE INFRASTRUCTURE DETECTION ALGORITHM

The chapter will describe and define the components of the context-aware DAS architecture according to the literature. These components are summarized in three subsystems, which are sensing subsystem, reasoning subsystem, and application subsystem. Based on these descriptions, a new layer, the context identification layer, will be proposed as additions to the reasoning subsystem. The structure of the context-aware system after this addition will be described. The proposed new layer is composed of two algorithms; namely, the infrastructure detection algorithm and the driver classification algorithm. This chapter will define and discuss the first algorithm and the use of deep neural networks to build a classifier for this algorithm using the extracted events of interest from the previous chapter. Moreover, the training and testing results of the developed neural network and the influence of the behavioral measures as input variables on the accuracy of the neural network will be discussed.

4.1 CONTEXT-AWARE DAS ARCHITECTURE

Context-aware systems are defined as the systems that continuously perceive, interpret, and react to the changing driving situations and the variation in the road conditions (Vahdat-Nejad et al., 2016). Such systems consist of three main subsystems, namely, sensing subsystem, reasoning subsystem, and application subsystem (Al-Sultan et al., 2013; Böhmländer et al., 2017; Loke, 2006; Vahdat-Nejad et al., 2016). These layers are responsible for the contextual information flow, analysis, and decision-making within the system.

The first subsystem is the sensing subsystem which collects the relevant contextual information regarding the host vehicle, the surrounding vehicles, and the surrounding environment. The sensing subsystem is usually equipped with a group of radar and LiDAR sensors, video camera(s), navigation system (e.g., GPS), host vehicle motion data collection tool (e.g., CAN bus or speed and accelerometers sensors) (Al-Sultan et al., 2013; Böhmländer et al., 2017). The sensing subsystem usually contains an acquisition data unit to fuse the raw data collected from different sources (e.g., sensors, videos, etc.). Moreover, the sensing subsystem will perform some preliminary analysis on the raw data and prepare and abstract the low-level data in a proper format for the next layer (i.e., reasoning subsystem) (Al-Sultan et al., 2013; Vahdat-Nejad et al., 2016). In some context-aware systems, a separate acquisition layer is added to the system for the same purpose and, in this case, the sensing subsystem will have two separate layers. The first one is the sensing layer which will be limited to data collection and the second layer is the acquisition layer (Böhmländer et al., 2017).

The second subsystem of the context-aware system is the reasoning or the thinking subsystem. This subsystem host layer(s) that is responsible for the decision-making process, which is the core of the context-aware system. Also, this subsystem processes the collected raw data to bridge between the collected low-level details and the required high-level details that are relevant to the context (Vahdat-Nejad et al., 2016). The reasoning subsystem includes a reasoning layer and a storage layer. The reasoning layer includes all the necessary algorithms that decide on the system actions (e.g., braking, warnings, or airbags activation), which are controlled by a control unit. These algorithms vary with the purpose of the context-aware DAS. For example, if the system is developed to detect the status of the driver (i.e., drunk, fatigue, etc.), the algorithms will be for driver behavior detection (Al-Sultan et al., 2013). The other layer (i.e., storage layer) stores all the relevant context and driver information such as road maps, relevant crash data, and driver actions history (Al-Sultan et al., 2013; Böhmländer et al., 2017).

Finally, the third subsystem in the context-aware systems is the application system, which is the acting subsystem. This subsystem translates the decision made by the reasoning subsystem into actions such as warning message to the driver or the surrounding drivers, activation of safety actuators (e.g., airbags), provide certain service to the driver (e.g., available parking spots), etc. (Al-Sultan et al., 2013; Böhmländer et al., 2017; Vahdat-Nejad et al., 2016).

The above-discussed context-aware system architecture is adopted in this dissertation since it is a more comprehensive architecture. However, there are several context-aware architectures that were proposed throughout the literature and were limited to a specific application only. For instance, a four-subsystems context-aware system was proposed to assess driver behavior using smartphone sensors data (Bejani and Ghatee, 2018). These four subsystems were smartphone calibration subsystem, feature clustering subsystem, context awareness subsystem, decision fusion subsystem. These proposed four subsystems were put together to process the smartphone data and to decide on the driver behavior either dangerous or normal (Bejani and Ghatee, 2018). Furthermore, a three-component context-aware system was proposed to estimate travel time based on taxis GPS trajectory data (Tang et al., 2018). These three components were map matching, travel time tensor construction, and travel time tensor factorization. The framework of the system including the three components was designed to serve the travel time estimation only (Tang et al., 2018). Moreover, Satzoda and Trivedi (2015) proposed a framework to enhance the lane changing applications. This framework consisted of two modules, namely, context definition module and lane detection module.

4.2 THE PROPOSED CONTEXT-AWARE ARCHITECTURE

As discussed in the previous section, this dissertation adopts a context-aware architecture that consists of three subsystems; sensing, reasoning, and application subsystems. As discussed in

Chapter Three, there was an evident difference between the driver behavior at the proximity of the intersections and on segments. Thus, a specific DAS that is designed based on the driver behavior as a whole without considering the effect of the surrounding infrastructure (e.g., intersections) will not be necessarily accepted by the drivers at all the network locations. For instance, a typical FCW sent to the driver while driving on a segment could be an early warning for the same driver when driving near an intersection. Same could happen if the DAS was designed based on the behavior of the normal drivers only without considering other driving classes such as cautious or aggressive drivers. Again, a typical FCW sent to a normal driver could be a late warning for a cautious driver. Therefore, the goal is to amend the reasoning subsystem to enhance the adaptability of the entire system to both (i) the driver's behavioral variation, and (ii) the alteration of the surrounding environment. These two amendments together were not adequately addressed in the literature as discussed in chapter two.

The proposed amendment on the reasoning subsystem is summarized in Figure 16. This amendment is represented as an additional layer (i.e., context identification layer) before the reasoning layer, which has the core algorithms of the context-aware DAS. The context identification layer uses algorithms to enhance the entire system adaptability to any alteration in the driving context. The context identification layer consists of two algorithms, namely, infrastructure detection algorithm, and driver classification algorithm.

The infrastructure detection algorithm is responsible for detecting when the driver is affected by driving at or in close proximity to an intersection. Hence, this algorithm could direct the system towards an intersection-specific application or a general (i.e., non-intersection) application. Moreover, this algorithm could help to customize and adjust the DAS parameters to either intersection- or general- specific parameters. In other words, in case the infrastructure detection algorithm identified that a driver is to be affected by an intersection at a certain location, the DAS will automatically apply the intersection parameters and make decisions based on these parameters. The intersection parameters could include, for example, shorter following distance or shorter TTC than the general system parameters. Furthermore, this algorithm feeds the detection output to the driver classification algorithm to identify the driver's class.

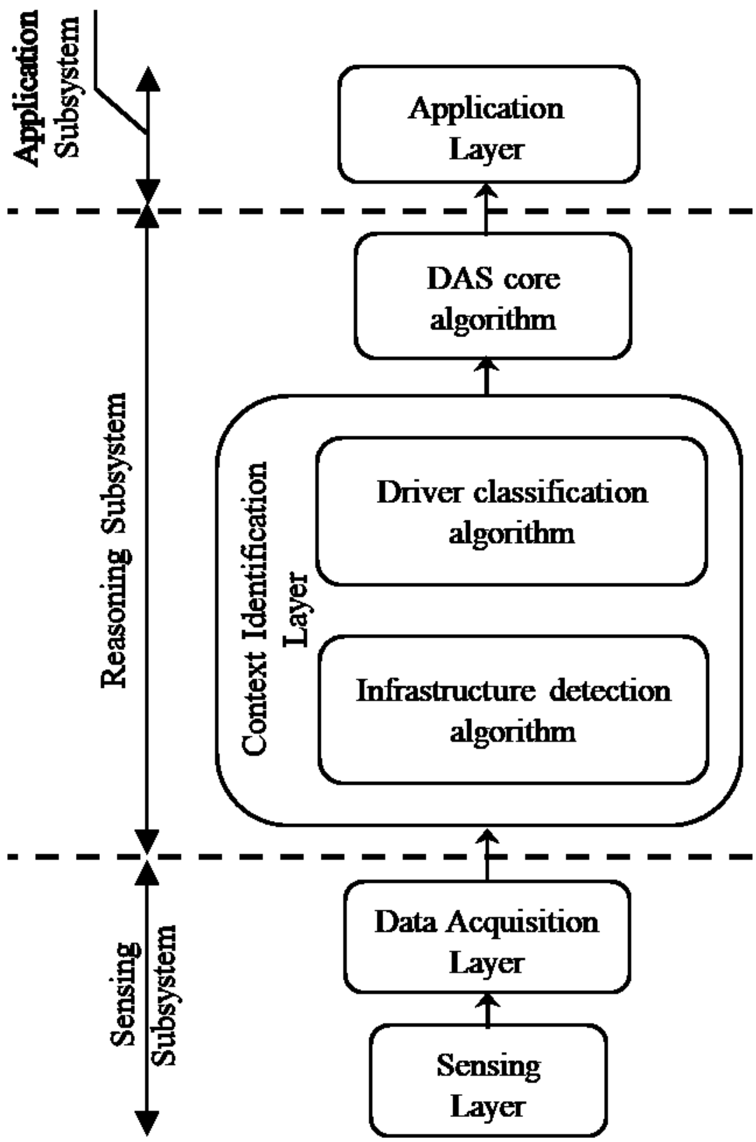


Figure 16. The proposed context-aware DAS architecture.

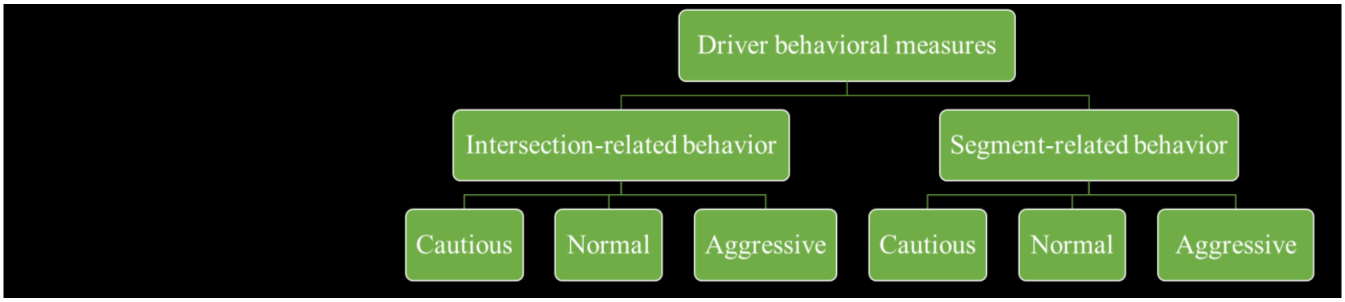


Figure 17. A summary of the context identification layer decisions.

On the other hand, the driver classification algorithm controls the identification of the drivers' classes (i.e., cautious, normal, or aggressive). The algorithm classifies the drivers according to their behavior and based on this, as discussed in the infrastructure detection algorithm, a group of customized parameters could be applied to the DAS. The customized parameters could be a longer following distance or a longer TTC for cautious drivers. For example, if a driver is classified as a cautious driver, a warning based on the customized parameters will be relayed to him/her earlier than the normal driver.

In summary, a context identification layer is proposed as an addition to the reasoning subsystem in a context-aware system architecture. The main purpose of this layer is to increase the adaptability of the system to the change in the driving context (i.e., the change in the surrounding infrastructure and the change in the driver behavior). To achieve this purpose, the infrastructure identification algorithm and the driver classification algorithm are defined as components of the context identification layer. These two algorithms will be discussed in detail in the following sections in this chapter and in the next chapter.

4.3 Neural Network Model for the Infrastructure Identification Layer

Neural networks are used as a deep learning technique to develop the infrastructure detection algorithm for the context identification layer. As discussed earlier, investigating the driver behavior on an individual level is crucial for the development of any assistance system because of the variation in the driving behavior and the different needs among the drivers (Bengler et al., 2014; Higgs and Abbas, 2015). Therefore, neural network models will be developed for each individual driver to accommodate the drivers' needs and preferences.

This section will discuss in brief the use of neural networks in driver behavior studies. Then, the training and validation of the individualized neural network infrastructure detection algorithm will be explained in detail. This explanation will include the input variables used for the neural network and the output variable. Moreover, the architecture of the neural network will also discuss including the model size, the activation function, and the optimization technique used to train the model.

4.3.1 Artificial Neural Networks and Driver Behavior

Artificial neural networks have been widely used as a modeling and pattern recognition technique in various transportation applications. For instance, neural networks were used to study the route choice behavior (Yang et al., 1993), traffic congestion prediction (Dougherty et al., 1993), operating speed predictions (McFadden et al., 2001), and traffic safety analysis (Abdelwahab and Abdel-Aty, 2002; Chang, 2005). Recently, with the revolution in communication technologies, neural networks showed superior performance over other machine learning techniques (e.g., support vector machine) learning transportation modes of travel using smartphone data (Fang et al., 2017). Moreover, the introduction of NDS, neural networks have been adopted to detect driver distraction and inattention (D'Orazio et al., 2007; Ye et al., 2017).

For driver behavior studies, neural networks were used to simulate driver behavior in a car-following situation and in safety-critical situations (Chong et al., 2013). The proposed neural network was trained using NDS data from trucks and cars. The variation among drivers was discussed in this study by calibrating the model using the data of one driver and then validating the same model using the data from a different driver. The results showed a drop in the model's validation accuracy when using data from a different driver, which claimed to be an indication for the heterogeneity among drivers (Chong et al., 2013).

Colombaroni and Fusco (2014) used different neural network architectures to model car following drivers' behavior. The results showed that a simple neural network with one hidden layer that consisted of five hidden units was an optimum neural network in terms of complexity and efficiency. The input variables of this network were relative speed, and the spacing between the following and leading vehicles and the output variable was the following vehicle acceleration. It was assumed that acceleration is the main measure that the driver can control directly through the braking and acceleration pedals. It is worth mentioning that a more complex neural network in terms of the number of input variables was also used. This complex neural network has six input variables including the relative speed and the spacing for time intervals of 1.5, 1, and 0.5 seconds before the incidents of interest (Colombaroni and Fusco, 2014). On the other hand, Khodayari et al. (2012) trained a neural network with one hidden layer with nine hidden units to predict the acceleration of the following vehicle. The input variables in this study were the relative speed, the distance between the leading and the following vehicles, the speed of the following vehicle, and the driver's instantaneous reaction delay (Khodayari et al., 2012).

In order to propose a safe curve speed model, Chen et al. (2012) trained backpropagation neural network using the driving data from 18 drivers. The neural network consists of one hidden layer only with 60 hidden units. Three main input variables were used for the training process including the speed (calculated based on the curvature, the friction coefficient, and roll angle), yaw rate, and adhesion workload (calculated based on the lateral and the longitudinal accelerations). The output variable was the predicted safe speed, which was compared to the speed that the drivers actually used (Chen et al., 2012).

For lane changing behavior prediction, Peng et al. (2015) collected driving behavioral measures from sixteen drivers during an on-road experiment to train a feedforward backpropagation neural network. The model was used to predict either a lane keeping or lane change maneuver which means that there was one output variable. The trained neural network had only one hidden layer that consists of fifteen hidden units. The input layer included seven variables, namely, vehicle speed, the use of turn signal, the distances and the time-to-collision between the host vehicle and the vehicles in front and behind in the current lane and in the target lane, and driver's head movement angle in the horizontal direction (Peng et al., 2015).

Moreover, (Meseguer et al., 2013) trained two backpropagation neural networks with one hidden layer that included nine hidden units to classify driver according to the road type (i.e., highway, suburban, and urban) where they were driving, and their behavior (e.g., normal, aggressive, and quiet). The input layer for both networks combines six variables, namely, the mean and the standard deviation for speed, acceleration, and the revolutions per minute of the vehicle's engine.

In summary, the neural networks have been extensively adopted for different aspects of driver behavior studies. For instance, neural networks were used to model and simulate car-following behavior based on the trajectories (Chong et al., 2013) or based on predicting specific measures such as acceleration (Colombaroni and Fusco, 2014; Khodayari et al., 2012). Also, this machine learning technique was used to model drivers' curve speed (Chen et al., 2012). Moreover, it was used as a classifier to predict drivers lane changing behavior (Peng et al., 2015) and to classify the drivers based on the road type (i.e., urban, suburban, highway) and their own behavior (i.e., quiet, normal, or aggressive) (Meseguer et al., 2013).

In general, the architecture of the neural network (i.e., number of hidden layers) is a tradeoff between the complexity of the network, which will affect the training time of the model and the targeted efficiency. Moreover, the effect of the input variables on the performance of the neural networks is not usually investigated. In contrast, the selection of the input and output variables was done in a pragmatic manner, for example, by selecting the acceleration as an output variable of a car-following model (Colombaroni and Fusco, 2014; Khodayari et al., 2012) or selecting them based on the problem definition such as in lane changing prediction (Peng et al., 2015).

4.3.2 Individualized Artificial Neural Networks for Driver Behavior Classification

The concept of the neural network modeling approach was initiated to mimic how the human brain works. Similar to the neuron in the human brain, which has a soma and an axon, each node (i.e., hidden unit) in the neural network has an input and an output (McCulloch and Pitts, 1943). These inputs and outputs connect all the nodes to represent the layered structured neural network exactly as in the nervous system where the neurons are connected by somas and axons. As shown in Figure 18, a deep neural network consists of an input layer, hidden layers, and an output layer.

For the infrastructure detection layer, a deep feed-forward neural network with backpropagation (Hecht-Nielsen, 1989) was chosen to detect whether the driver behavior was influenced by the proximity to an intersection. Neural networks with backpropagation showed high performance in pattern recognition and classification for transportation-related problems as discussed earlier in the previous section. The conjugate gradient algorithm was used as the optimization function (Charalambous, 1992). A *Log-Sigmoid* function was used as an activation function (i.e., transfer function) for each layer.

In order to reach the optimum detection performance (i.e., high accuracy with less training time), the neural network was trained with two hidden layers and three hidden layers. It is worth mentioning that when the network is deeper (i.e., has more hidden layers), the accuracy is expected to increase. Nevertheless, the deeper networks are trained slower than the shallower networks. Therefore, only two different network sizes will be checked (i.e., two and three hidden layers) to limit the scope of this analysis. Moreover, it is anticipated that the wider the network is, the higher the performance will be. Similar to the deeper networks, the wider networks (i.e., the networks with more hidden units in each hidden layer) are expensive in terms of the training time. Hence, a different number of hidden units were used for each hidden layer (i.e., from 5 to 60 hidden units with an increment of 5 hidden units) to find the optimum accuracy for the model.

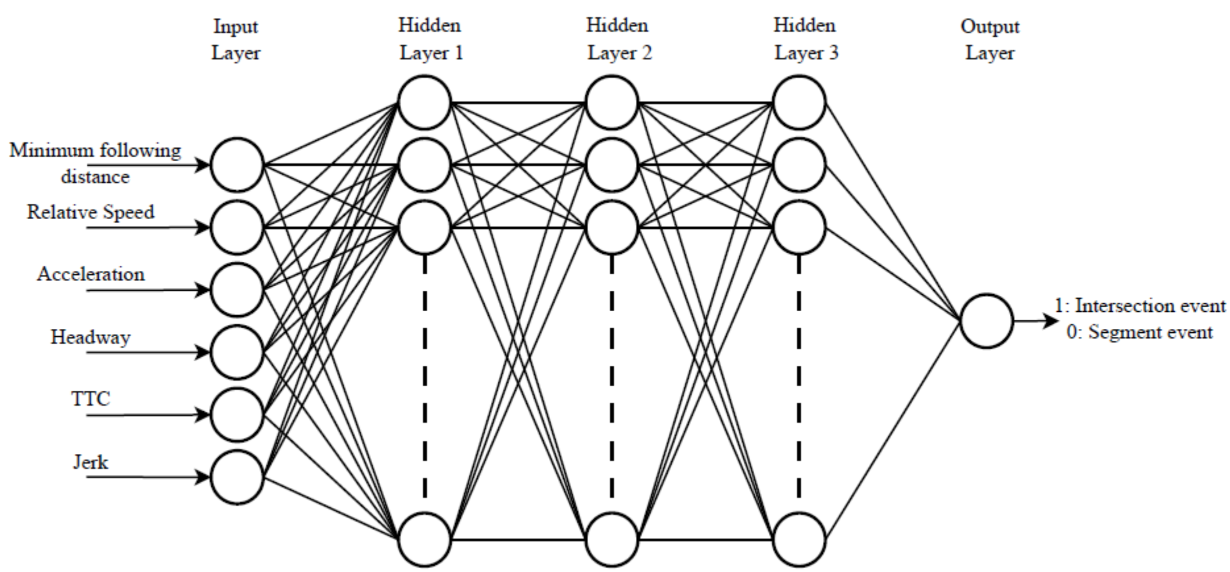


Figure 18. An example of the 3-hidden layers neural network model structure.

The above-described neural network architecture was trained for each individual driver, which means that each driver had a customized model that classifies his or her behavior into intersection or segment-related behavior. The input variables (i.e., features) of the input layer for these networks were the driver behavioral measures, which were extracted from the NDS as discussed in Chapter Three. The input variables were minimum following distance, deceleration, jerk, relative speed between the host and the leading vehicle, time headway, and TTC. These input variables were extracted for each training example (i.e., incidents of interest) and combined in a $m \times n$ data (i.e., input) matrix Xi. Figure 19 shows a summary of the total number of the training examples that were extracted for each driver. In addition, the output variable (i.e., class) that was obtained from the mapping and classification step was represented by a $m \times 1$ matrix $Y_{i,}$ which is called the target matrix. Both matrices (i.e., X_i and Y_i) were represented below.

$$X_i = \begin{bmatrix} x_1^{(1)} & \cdots & x_n^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(m)} & \cdots & x_n^{(m)} \end{bmatrix}, Y_i = \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{bmatrix}$$

where X_i and Y_i are respectively the input and target (output) matrices of driver i, $x_n^{(m)}$ is the variable n for the training example m, and $y^{(m)}$ is the target class of a training example m.

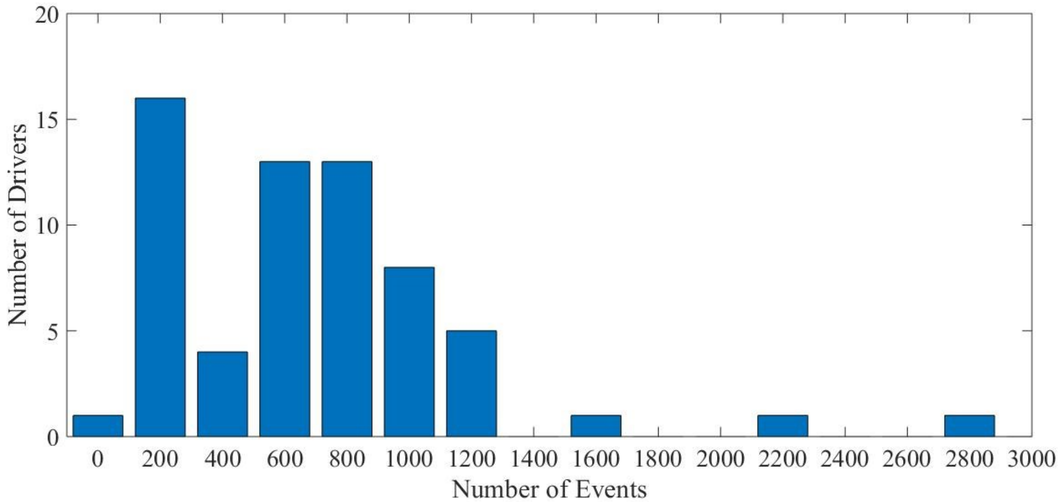


Figure 19. The frequency of the drivers with respect to the number of the events.

The values of the target class from the output layer are either "1" for intersection events or "0" for midblock segment events as shown in Figure 18. It is worth mentioning that the number of training examples of each driver varies based on the extracted incidents of interest for each driver, and the total number of features for each training example is six (i.e., the number of the behavioral measures) as shown in Figure 18. Since each variable in the input matrix had a wide range of values, the input matrix was normalized to have all the variables with zero mean and a variance of one. The primary goal of the normalization step is to enhance the training process of the neural network model. The mean normalization was carried out based on the following equation.

$$\bar{X}_i = \frac{X_i - \mu_i}{\sigma_i} \qquad (4)$$

where $\bar{X}_i$ is the normalized input (i.e., feature) matrix for driver i, and μ_i and σ_i are the mean and the standard deviation respectively of an input matrix X_i for each driver i.

To train this network, the extracted incidents of interest of each driver were split randomly to 70% as a training set, and 30% as a test set (Fang et al., 2017; Khodayari et al., 2012).

4.4 SELECTING THE OPTIMUM NEURAL NETWORK SIZE

Deep neural networks with varied sizes were trained to find the optimum neural network size for each driver's model. The optimum size is defined in the context of this dissertation as the neural network size, which provides the highest possible classification accuracy with the least training time. This optimum size will ensure the best performance of the infrastructure detection algorithm as part of the reasoning subsystem, so it will not significantly affect the performance of the context-aware DAS as a whole. The primary purpose of the model was to classify drivers' car-following events while braking based on how each driver's behavior could change due to the driving context (i.e., the interactions with surrounding vehicles and the proximity of intersections). Moreover, the models were developed for each individual driver to accommodate the different preferences and needs for each driver. The accuracy of each neural network model was calculated as the percentage of the number of the events, which were correctly classified by the model for each driver. Then, for each network size, the average total, training, and test accuracies were calculated for all the drivers and plotted in

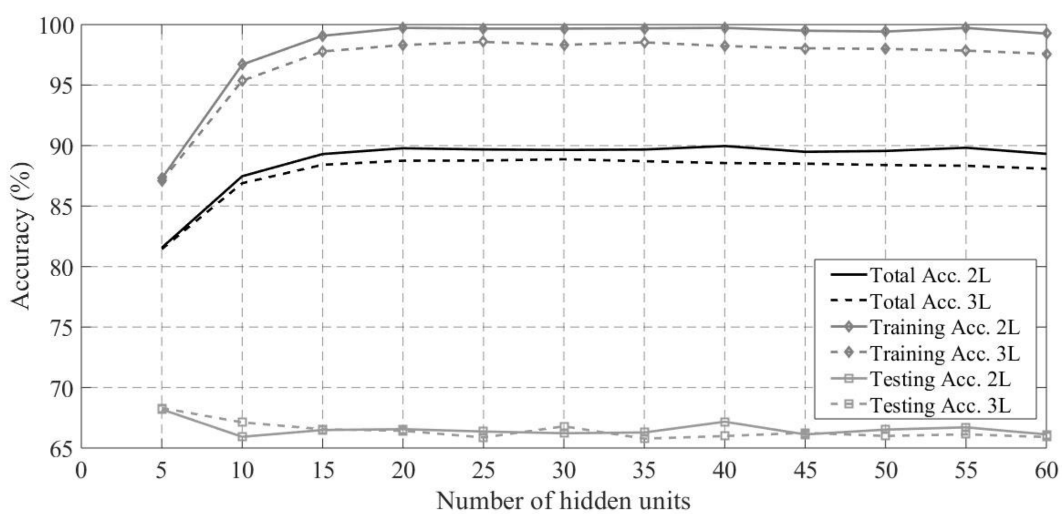


Figure ***20***.

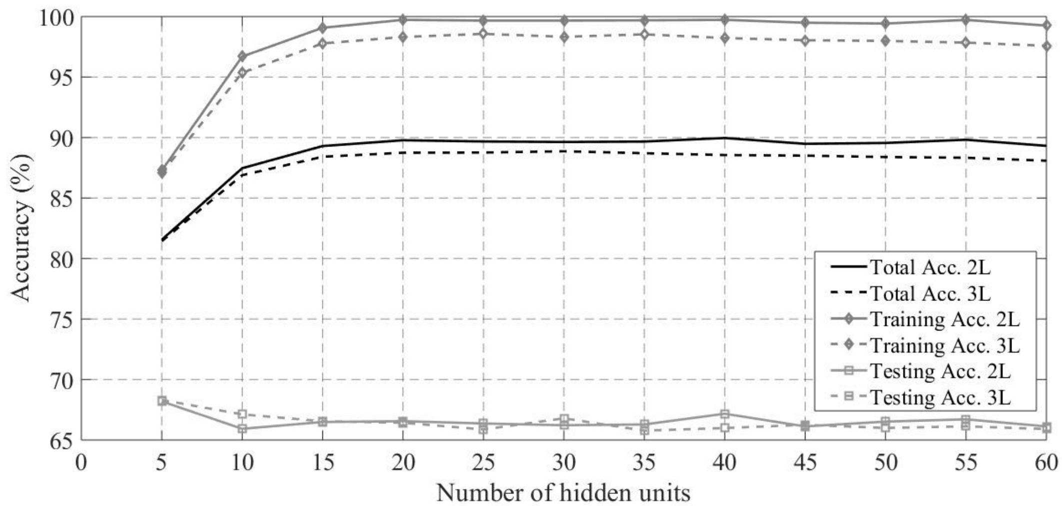


Figure ***20*** shows that the accuracies (i.e., training, test, and total) of the neural network model varied based on the network size (i.e., number of hidden layers and hidden units). As shown in the figure, the training accuracy was very high, and the test accuracy was relatively low yet reasonable. Furthermore, the size did not have a significant effect on the accuracy after exceeding a specific network size. For example, for two hidden layers networks, the average accuracies of the networks, which had 25 hidden units in each layer or more were almost the same. Furthermore, there was no significant difference (at 5% significance level) between the accuracies of using either two (2L) or three (3L) hidden layers neural network models. Nevertheless, the training time increased significantly with the increase in the number of hidden layers and/or the hidden units.

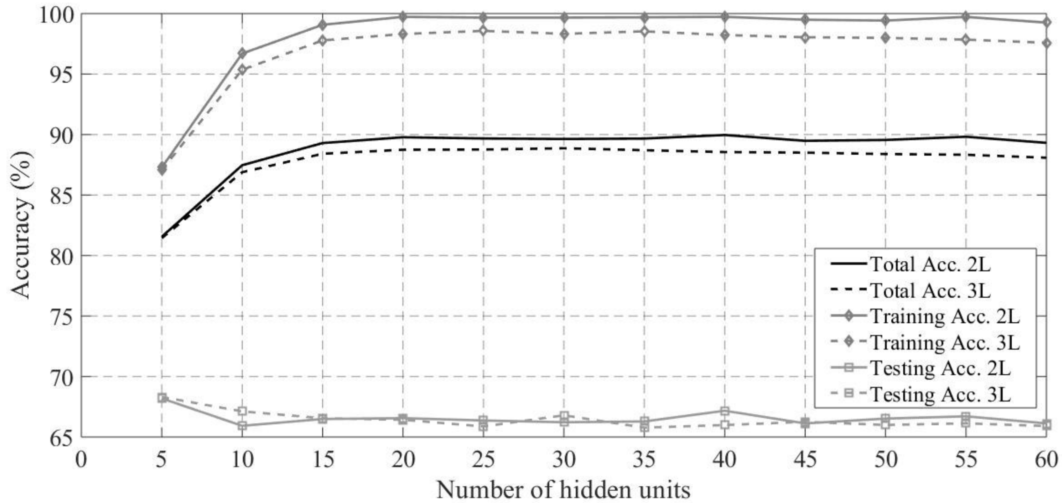


Figure 20. The average classification accuracy (Acc.) of the 63 drivers for each neural network size.

Figure 21 shows an example of the confusion matrix as a result of driver behavior classification of one of the drivers. A confusion matrix shows the relation between the target (i.e., actual) and the output (i.e., predicted) classes. The target and the output classes are shown on the horizontal and the vertical direction, respectively. When the results of the output class match the target class, then, the classification of the model is correct. The bottom right corner of each matrix has two numbers. The upper number is the percentage of the correct classifications (i.e., accuracy) and the lower number is the percentage of the error classifications. Figure 21 includes six matrices for six different neural network sizes. These six sizes are 5, 25, 60 hidden units for two and three hidden layers neural network for the same driver.

As shown in Figure 21, the differences between the accuracies of the two and three hidden layers were minimal and insignificant. For instance, the classification accuracy of the network, which had two hidden layers with five hidden units each, was 86.3%, which was only 0.9% higher than the accuracy of three hidden layers each with five hidden units. However, the training time increased by around 23%. The same trend was also observed for the networks with 25 and 60 hidden units where the training time increased substantially while there was no significant difference in the accuracy as shown in the same figure.

Figure 21. An example of the classification results for different neural network sizes.

On the other hand, when comparing the classification accuracy of the five hidden unit networks and the 25 hidden unit networks, the accuracy was considerably increased (i.e., from 86.3% to 92.7%). Conversely, the classification accuracies of the 25 and 60 hidden unit networks were almost the same (i.e., 92.7% and 92.3% respectively). Again, the training time also increased by about 60% when the hidden units were increased. The discussed trend and results of this driver were observed for all the other 62 drivers. Therefore, a neural network of two hidden layers with 25 hidden units each was considered as the optimum neural network size in terms of accuracy and training time to detect the change in driving behavior near intersections. The average total accuracy of this network for all drivers was 89.8% with an average of 99.7% training accuracy and an average of 66.7% test accuracy. Figure 22 and Figure 23 summarize the total, training, and test accuracies of the above-mentioned neural network for all the drivers described as the frequency of the drivers in each accuracy bin. As shown in the figures, the models had a reasonable total accuracy and significantly high training accuracies. However, a wide range of variation was observed for the test accuracy. The reason for this variation was due to the variation of the available number of events. For example, the test accuracy was significantly low for some drivers due to the availability of a small number of events (i.e., training examples) for model training and test for those drivers.

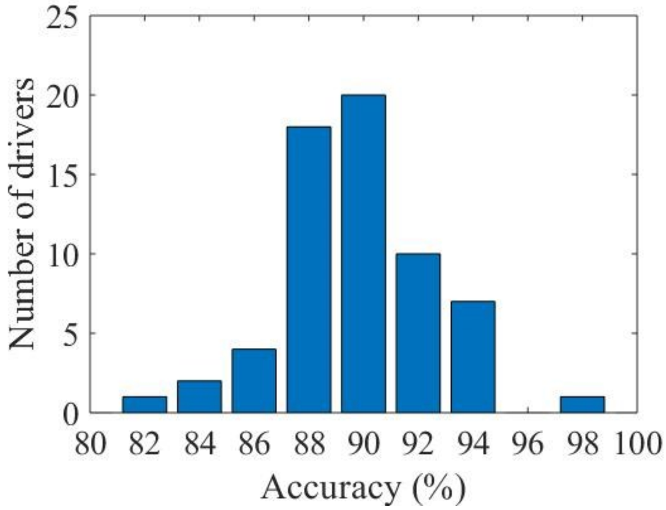


Figure 22. Summary of the total accuracies for all the drivers.

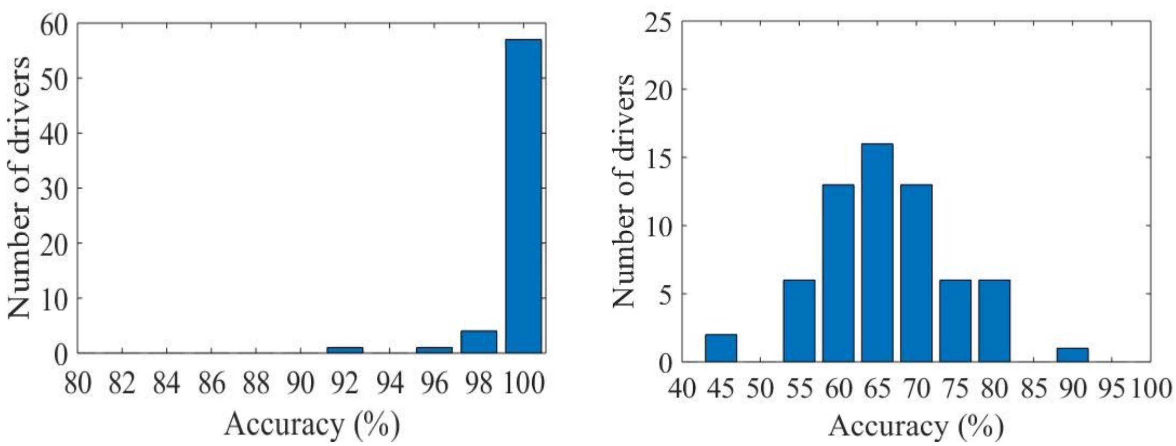


Figure 23. Summary of the training (on the left) and test (on the right) accuracies for all the drivers.

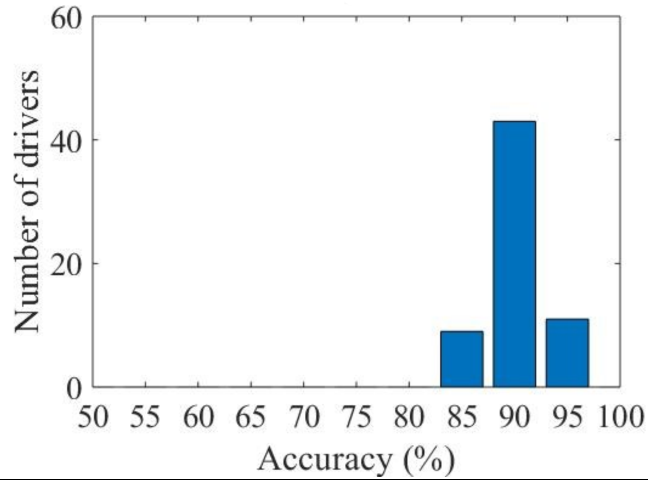
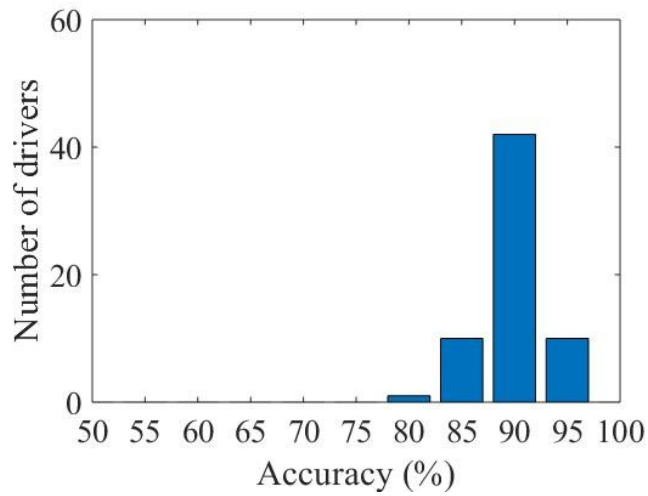
Despite the high accuracy of the neural network models, it is not clear from these models which variable of the behavioral measures had a significant impact on the results. Thus, a sensitivity analysis will be performed in the next section to find, which variables could significantly affect the accuracy of the model.

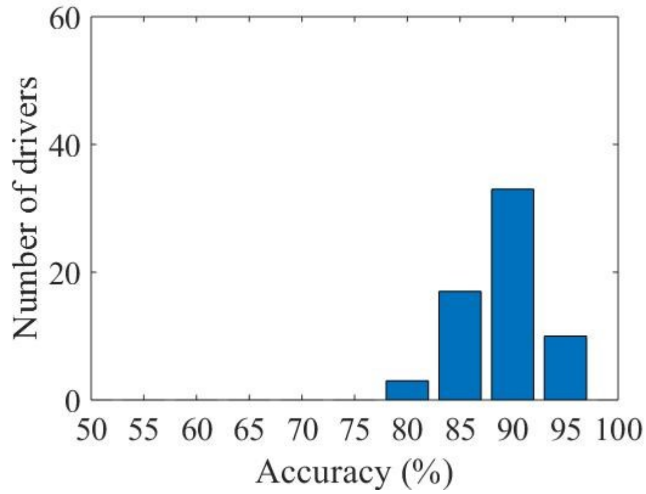
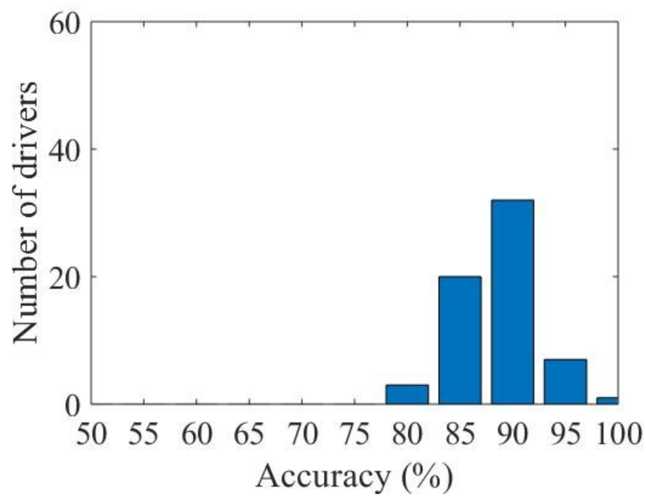
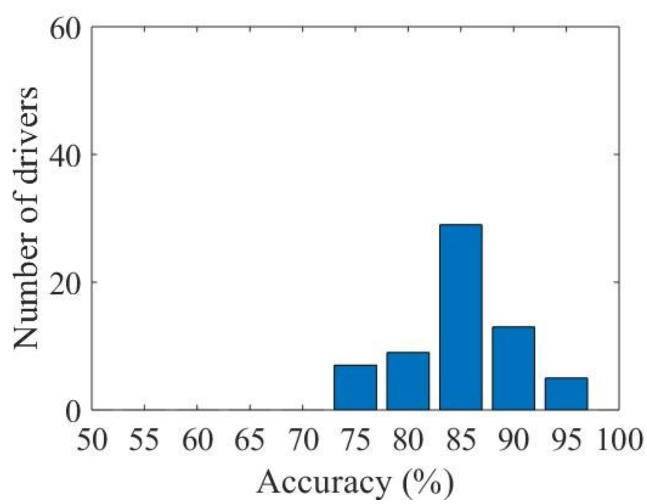
4.5 THE INFLUENCE OF INPUT VARIABLES ON THE NEURAL NETWORK MODELS

For the sensitivity analysis, the selected neural network, that consists of two hidden layers and 25 hidden units for each layer, was trained several times with different random combinations of input variables. The main goal behind this process is to discover which combination of the input variables will have a significant effect on the classification accuracy. In other words, this process aims to find which behavioral measure(s) will be substantially affected by the proximity to the intersection.

Table 4 summarizes the results of the sensitivity analysis described by the frequency of the drivers in each total accuracy bin. This table summarizes eleven combinations between the input variable, including the combination discussed-above, and the corresponding average total accuracy. The results in Table 4 were ranked in a descending order starting from the highest average total accuracy. As shown from the table, the top accuracy was obtained by including all the six input variables (i.e., behavioral measures). Moreover, as shown from the frequency of this case, all the drivers had high classification accuracy (i.e., at least a total classification accuracy of 80%).

Table 4 *Total accuracies for different input variables combinations*

No.	Input Variables	Average Total Accuracy (%)	Accuracy Frequency for All Drivers
1	Following distance, acceleration, relative speed, headway, TTC, jerk	89.8	
2	Following distance, acceleration, relative speed, headway, jerk	89.6	
Table continues in the next page >>			

No.	Input Variables	Average Total Accuracy (%)	Accuracy Frequency for All Drivers
3	Following distance, acceleration, relative speed, headway	88.9	
4	Following distance, acceleration, relative speed, headway, TTC	88.7	
5	Following distance, acceleration, relative speed, TTC	85.1	
Table continues in the next page >>			

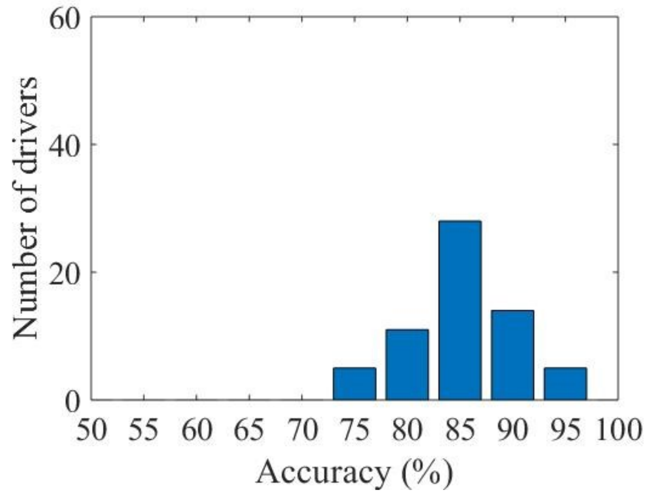
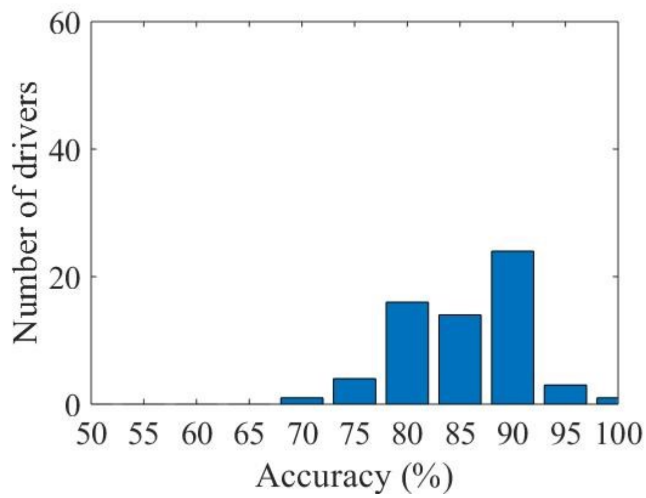
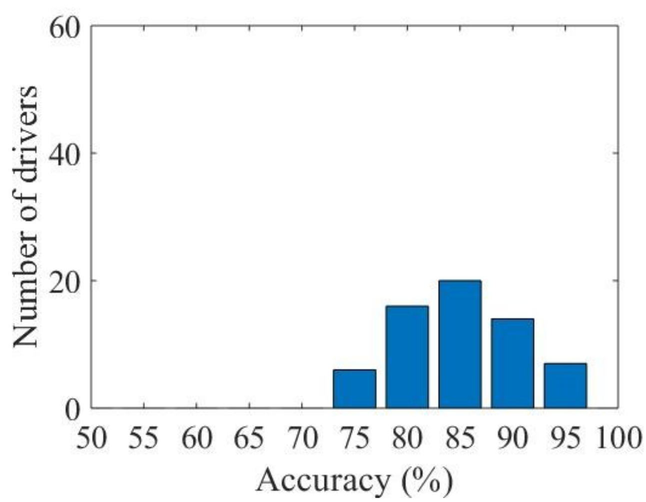
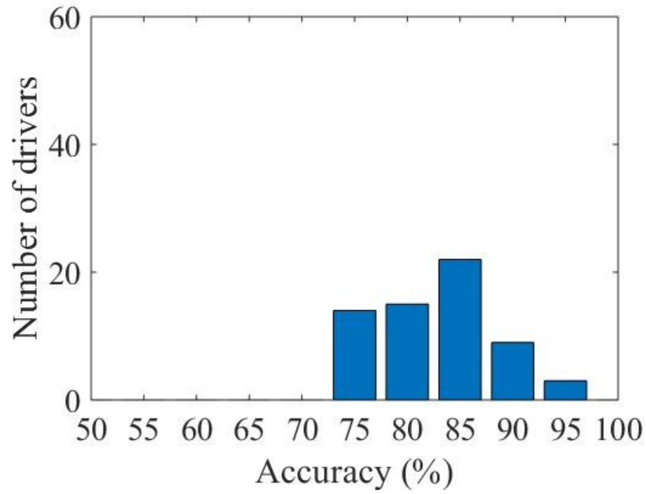
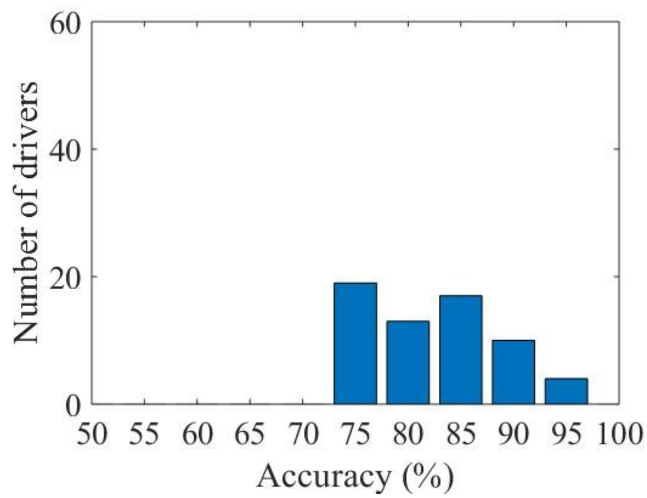
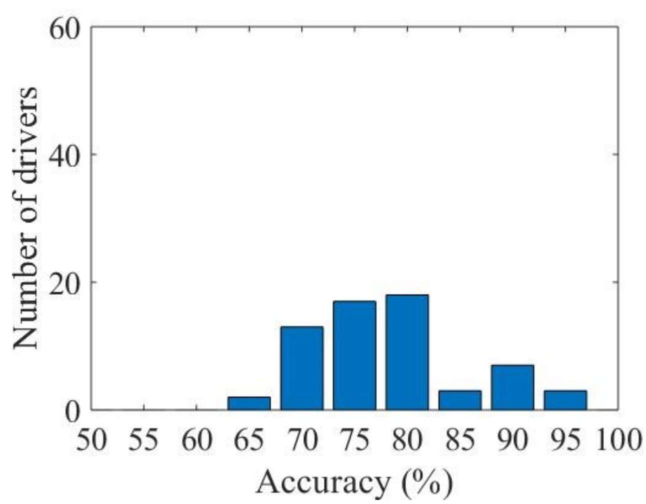
No.	Input Variables	Average Total Accuracy (%)	Accuracy Frequency for All Drivers
6	Following distance, acceleration, relative speed	85.1	
7	Following distance, acceleration, headway	85.1	
8	Following distance, relative speed, headway	85.1	

Table continues in the next page >>

No.	Input Variables	Average Total Accuracy (%)	Accuracy Frequency for All Drivers
9	Following distance, acceleration, TTC	82.7	
10	Acceleration, relative speed, TTC	82.5	
11	Following distance, relative speed, TTC	78.4	

The first four input variables combinations in Table 4 showed that both TTC and jerk did not have a significant effect on the classification accuracy and the frequency distribution of the drivers did not have a substantial change. For example, when excluding the TTC from the input variables (as in combination number 2), the average total accuracy decreased only by 0.2% and the comparison of the frequency distributions of the total accuracy showed insignificant change at 95% confidence level. Also, when comparing the combinations number 5 and 6, the exclusion of the TTC had no effect on the average total accuracy, and the frequency distributions of the accuracy for both cases were almost the same.

Moreover, excluding the jerk only form the input variables (as in combination number 4) resulted in a 1.2% drop in the average total accuracy. There was no significant difference between the frequency distribution of this result and the result of including all the variables as input variables for the network. Furthermore, the average total accuracy was 88.9% (dropped only 1%) when excluding the TTC and the jerk from the input variables. This drop was also insignificant at 95% confidence level when comparing the frequency distribution of the total accuracy. These observations indicated that the TTC and the jerk could be excluded while maintaining almost the same accuracy (i.e., performance). In other words, the TTC and the jerk as behavioral measures will not substantially affect the detection of when an individual driver is influenced by the proximity to an intersection.

On the other hand, when comparing combinations number 4 and 5, the exclusion of headway caused approximately a 3.6% decrease in the average total accuracy. This decrease reflected on the frequency distribution of the accuracy which changed significantly (at 95% confidence level). The same trend of accuracy deterioration was noticed when excluding the relative speed, the following distance, and the acceleration in combinations number 9, 10, and 11, respectively and comparing the results with combination number 5. For instance, the elimination of the relative speed from the input variables of the individual drivers' models reduced the average total accuracy of these models by 2.4%. Moreover, the elimination of the following distance and the acceleration in combinations number 10 and 11 decreased the average total accuracy by 2.6% and 6.7%, respectively. These results revealed that the headway, the following distance, the relative speed, and the acceleration could significantly affect the efficiency of the classification model. Hence, these behavioral measures could be the most impacted behavioral measures when a driver is approaching an intersection.

In addition, the combinations of input variables number 6, 7, and 8 had almost the same average total accuracy, and there were no significant differences between their accuracy frequency distributions. This result along with the above-discussed results indicated that at least three of the four variables (i.e., the following distance, the headway, the acceleration, and the relative speed) should be included in the input variables to obtain high accuracies. However, the results highlighted that the elimination of the acceleration form the input variables had the most substantial effect on the accuracy (i.e., 6.7% drop in accuracy) when compared with the other four variables. This observation revealed that the acceleration was the most influenced behavioral measure due to the proximity to intersections. Nevertheless, this conclusion should be handled with caution since the results depended only on the accuracies of the trained sixty-three models on an aggregate level. Moreover, one of the disadvantages of the neural networks is that it does not provide a clear relationship between the input and the output variables.

In general, the results presented in this section indicated that the used behavioral measures combinations were significant in detecting the change in the driver behavior at different locations on the road network (i.e., at the proximity to intersections). However, the TTC and the jerk do not affect this detection for individual drivers.

4.6 Summary and Discussion

This chapter emphasized the development of the infrastructure detection algorithm, which was proposed as one of two algorithms in the context identification layer. The context identification layer belonged to the reasoning subsystem, which is responsible for the decision-making process in the context-aware DAS. The infrastructure detection algorithm aimed at detecting the intersection-related driving behavior where the driver adjusts his/her behavior due to the presence of an intersection ahead of him/her. Moreover, this algorithm supported the purpose of the context-aware DAS in assisting drivers by fulfilling their needs and meeting their expectations at the above-mentioned situations on the road network. The output of this algorithm will be either "1" for a detected intersection-related behavior or "0" for a detected segment-related behavior. This output was proposed as the input for the next algorithm (i.e., driver classification algorithm) in the context identification layer.

Furthermore, this algorithm could be used alone to direct the information flow within the context-aware DAS to intersection or segment applications. The proposed algorithm could be a supplement to the context-aware DAS messaging in a V2I CV environment by identifying when the driver will be influenced by an intersection. For example, when a driver is approaching an intersection that supports the V2I communication, a warning message might be sent to the driver once the vehicle is within the communication coverage. This message could be relayed to the driver early or late since the driver could be influenced by the intersection before or after getting in the intersection coverage area. Therefore, the proposed infrastructure detection algorithm could adjust this warning message to the time when the driver is influenced by the intersection.

In order to develop the infrastructure detection algorithm, deep neural networks were trained for each driver using the extracted incidents of interest as training examples and the six behavioral measures as input variables. These six behavioral measures included the host vehicle acceleration, jerk, minimum following distance, relative speed, TTC, and headway between the host vehicle and the identified leading vehicle. Based on the perception of the surrounding infrastructure, the trained networks classify the driver behavior to an intersection or midblock segment behavior. An event of interest was defined as an intersection-related event based on the stopping sight distance computation as discussed in the third chapter.

Several neural networks with different sizes (i.e., 24 sizes in total) were developed and trained for each individual driver to find the optimal network size, which gives the best accuracy for detecting drivers' behavioral change due to their proximity to an intersection. These neural networks varied in depth (i.e., two and three hidden layers) and width (i.e., increased from five to 60 hidden units by increments of five units for each layer). The training and testing results of the neural networks showed that there was no statistically significant difference between the two hidden layers and three hidden layers neural networks. Among the trained neural networks, the optimum size that had the minimum training time and the maximum accuracy was found to be the two hidden layers network with 25 hidden units. This neural network could detect with high total accuracy (i.e., about 90% classification accuracy) whether the driver behavior was influenced due to driving near an intersection.

A sensitivity analysis was then carried out on the optimal neural network size to examine the effect of the input variables on the overall accuracy of the infrastructure detection algorithm (i.e., the average total accuracy of the sixty-three drivers). In this analysis, the neural networks were re-trained several times with ten different combinations of the input variables. The results of this indicated that the TTC and the jerk could be excluded from the input variables without substantially effecting the accuracy. On the other hand, the host vehicle acceleration, following distance, headway, and the relative speed could affect the detection accuracy significantly. Moreover, the host vehicle acceleration had the highest impact on the accuracy when eliminated from the input variables. In general, at least three of the four measures (i.e., the host vehicle acceleration, following distance, headway, and the relative speed) should be included in the input variables to reach a reasonable accuracy (i.e., 85% or above).

5. THE LOCATION-BASED DRIVER BEHAVIOR CLASSIFICATION ALGORITHM

This chapter will present the driver behavior classification algorithm as part of the proposed context identification layer in the context-aware DAS. In order to emphasize the significance of this algorithm, this chapter will discuss the differences that could arise from classifying the drivers based on their location (i.e., at intersections and on segments) into caution, normal, and aggressive drivers. An unsupervised machine learning technique will be discussed for the drivers' classification. The impact of the behavioral measures on the classification will be discussed as well.

5.1 DRIVER BEHAVIOR CLASSES

As discussed in the literature review chapter, previous studies proposed various definitions to describe the variation in drivers' behavior. For instance, some studies in the literature focused mainly on detection of a specific behavior class (i.e., aggressive or risky driving detection) (Castignani et al., 2015; Feng et al., 2017; Fitzpatrick et al., 2017; Gilman et al., 2015; Jahangiri et al., 2018; Johnson and Trivedi, 2011; Li et al., 2016; Rodriguez Gonzalez et al., 2014; Singh et al., 2017). The methodologies in most of these studies were based on driver behavioral data and the detection was constructed by investigating both normal and aggressive driving. However, these studies proposed the aggressive behavior detection based on certain thresholds of the behavioral measures regardless the differences between different behavior classes. On the other hand, some studies focused on drivers' classification based on comparing different behavior classes relatively. These studies included classifying the drivers into different behavioral classes as shown in Table

5, which summarizes previous studies that were concerned with driver behavior classification and the corresponding defined behavior classes.

Table 5 *Driver behavior classes summary*

Name	Defined Behavior Classes
(Wang et al., 2010a)	Risk prone and risk infrequent, stable and unstable, and non-skillful and skillful
(Casucci et al., 2010; Wang et al., 2010a)	Aggressive and prudent
(Rong et al., 2011)	Aggressive, moderate, and conservative
(Sundbom et al., 2013; Wang et al., 2017)	Aggressive and normal
(Hill et al., 2015)	Very conservative, somewhat conservative, somewhat aggressive, and very aggressive
(Chandrasiri et al., 2016, 2012)	High and low/average skillful
(Wu et al., 2016)	Minimal, slight, moderate, serious, and severe speeding prone and acceleration/deceleration
(Zheng et al., 2017)	Four types (A, B, C, and D)
(Osafune et al., 2017)	Safe and risky
(Yang et al., 2018)	Normal, aggressive stable, aggressive unstable, unaggressive stable, and unaggressive unstable
(Eftekhari and Ghatee, 2018)	Safe, semi-aggressive, and aggressive
(Bejani and Ghatee, 2018)	Dangerous and normal
(Fugiglando et al., 2018)	Various classes based on the classification measures

Although the classes in this table appear to be similar, the definitions of each of these classes were different from one study to another. For instance, Wang et al. (2010a) defined aggressive drivers as the drivers who keep a shorter headway to the leading vehicles and used to approach the leading vehicles in shorter TTC when compared to other prudent drivers. On the other hand, Wang et al. (2017) defined aggressive driving based on the speed and the throttle position in the longitudinal direction only. The throttle position was considered to have a direct relation to drivers' lateral and longitudinal acceleration and deceleration. Based on these two measures, drivers with high speed and small throttle opening were classified as aggressive drivers. Furthermore, Yang et al. (2018) defined the typical-aggressive drivers as the drivers who have large mean values of acceleration and lateral deviation and small values of the distance to the leading vehicle and to the yellow line and braking frequency. Casucci et al. (2010) defined the sensation seekers (i.e., aggressive drivers) as the drivers who have high speed with more abrupt lane changes, acceleration, and braking maneuvers. Rong et al. (2011) described the aggressive drivers as those who had shorter driving experience, they tended to drive at higher speeds and make lane changes with smaller gaps, and their perception and cognition times were shorter. For freeways, Hill et al. (2015) used vehicle speed and the number of discretionary lane changes as indicators for aggressive drivers. An aggressive driver was concluded to have longer lane change time than a conservative driver.

Using smartphone sensor data, Eftekhari and Ghatee (2018) defined aggressive behavior based on four behavioral measures, which were lateral and longitudinal acceleration, angular velocity, and angular variation of the vehicle. Aggressive behavior was identified using a hybrid of discrete wavelet transform and adaptive neuro-fuzzy inference system. The results of this study showed that the proposed classification methodology could still give good accuracy when excluding the angular velocity from the inputs. On the other hand, Fugiglando et al. (2018) investigated the dissimilarities between drivers using CAN bus signals and proposed a methodology to find the best measure to cluster the drivers. However, the clusters generated from this methodology were not explained in terms of the behavior explanation for each class. In other words, this study did not have a definition of the clusters generated from the proposed methodology (Fugiglando et al., 2018).

In summary, the literature had various definitions of aggressive drivers with aggressive driving being classified based on a huge number of measures ranging from physiological signals (Yang et al., 2018) and perception and cognition abilities (Rong et al., 2011) to actual driver behavioral measures such as speed, acceleration, headway, etc. (Casucci et al., 2010; Eftekhari and Ghatee, 2018; Hill et al., 2015; Wang et al., 2010a, 2017). Although, aggressive driving was defined as driving the vehicle in a pushy or impatient manner as a result of interactions with the driving environment (Neuman et al., 2003), the effect of the change in driving environment was not fully addressed in the literature. In other words, the previous studies did not compare the classification of the drivers on different road network locations, such as at intersection and on roadway segments. Therefore, this chapter will highlight the importance of integrating this change in the context identification layer of the context-aware DAS by comparing the classification of the drivers at different locations on the road network.

5.2 DRIVERS CLASSIFICATION PROCEDURE

To develop the driver behavior classification algorithm for the context identification layer, the classification procedure discussed in this section was carried out. This procedure used an unsupervised machine learning technique, which includes two main steps: Principal Component Analysis (PCA), and clustering analysis (i.e., *K*-means). Although this was an unsupervised procedure, the classification will be performed at different levels, as shown in Figure 24 and Figure

25, (i.e., location and driver behavior classification) to verify the results and to find the dissimilarities between the drivers' groups. It is worth mentioning that the driver classification algorithm will be based on the behavioral classification.

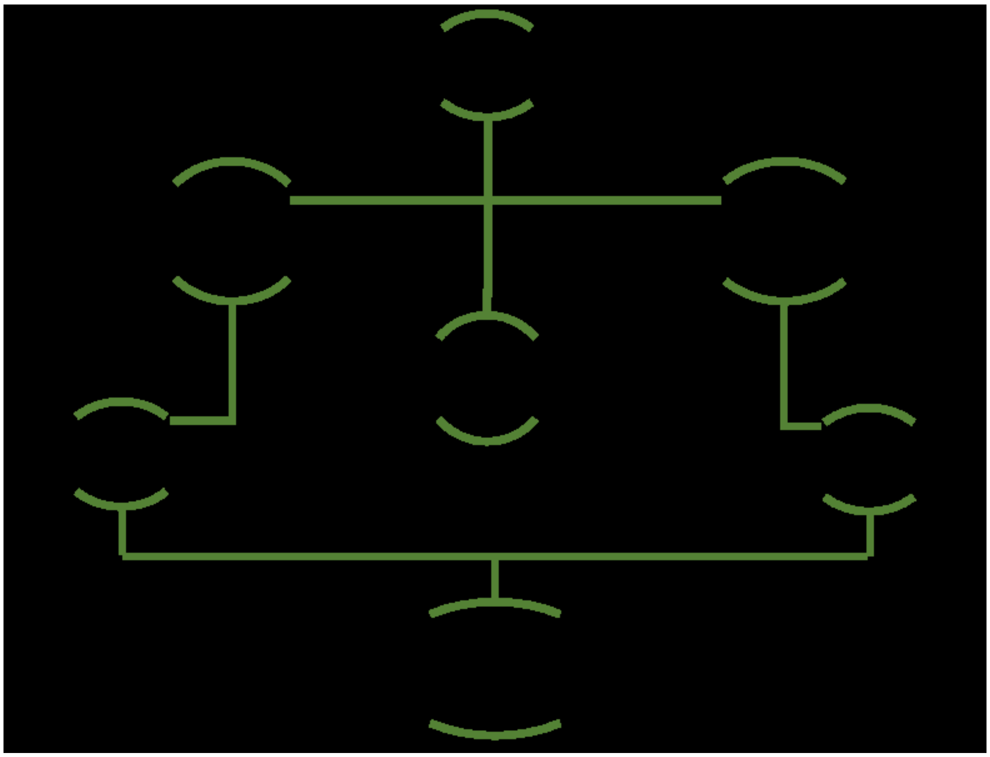


Figure 24. The aggregation of the extracted events.

As discussed earlier, a total of 44,383 events were extracted based on the procedure discussed in chapter two with an average of 704 events for each driver. These events were grouped and split, with the aid of ArcGIS, into intersection- or segment-related events. These events were used as training examples for the next step of the classification procedure. Each training example had six driving behavioral indicators (i.e., following distance, acceleration, relative speed, headway, TTC, and jerk). The training examples were aggregated on several levels and prepared for classification as shown in Figure 24 and Figure 25. First, the average of each behavioral measure was calculated for each driver in the aggregation step, and hence, the following matrix (X_{all}) was developed.

$$X_{all(63\times6)} = \begin{bmatrix} x_1^{(1)} & \cdots & x_6^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(63)} & \cdots & x_6^{(63)} \end{bmatrix}$$

where the variables x_1 to x_6 are the averages of the six driving behavioral indicators, and the matrix had 63 rows representing the 63 drivers.

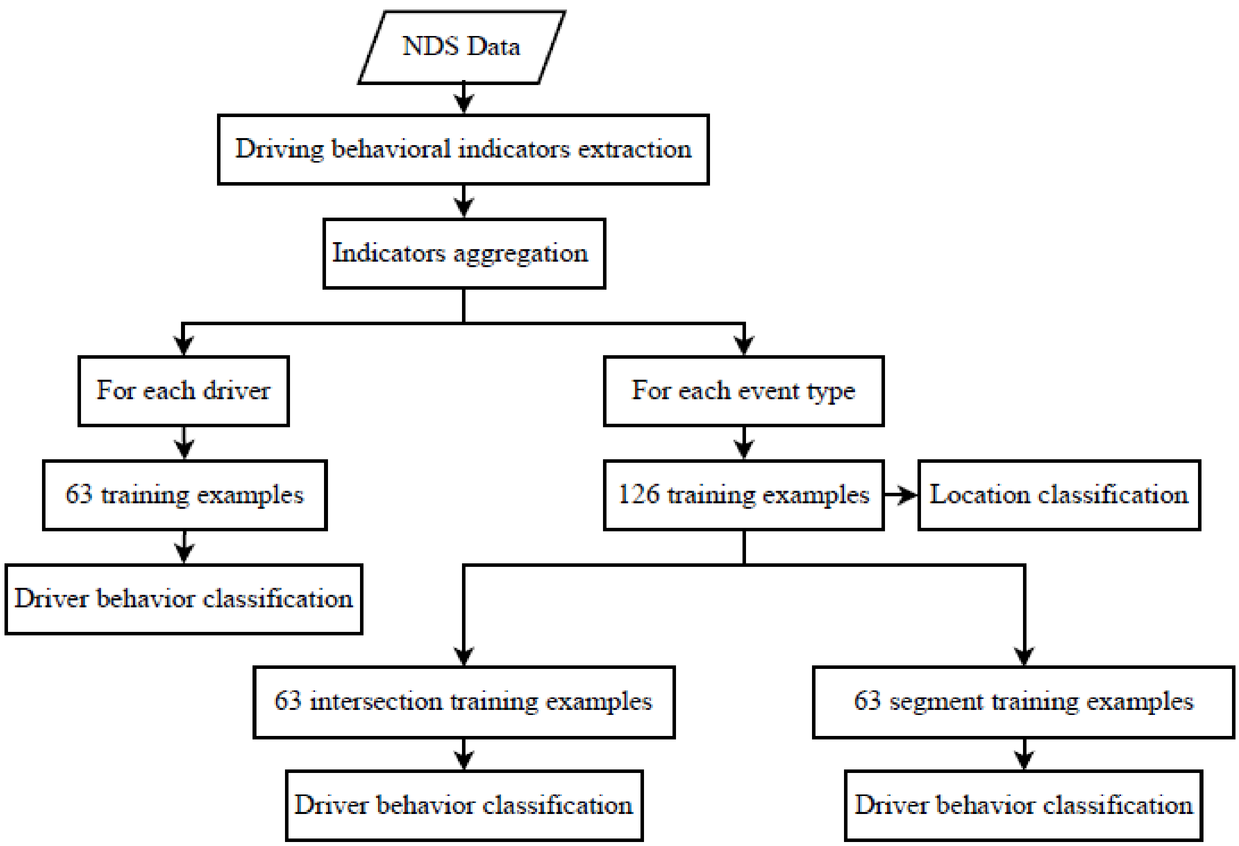


Figure 25. Events aggregation and driver classification framework.

Second, based on the events mapping, the average of each behavioral measure was calculated for each group (i.e., intersection- and segment-related events) and for each driver, as shown in Figure 24. In other words, each driver had two groups of indicators, namely, intersection and segment driving behavioral measures. This could be represented in the following matrices X_{int} and X_{seg} for intersection and segment events, respectively.

$$X_{int(63\times6)} = \begin{bmatrix} x_{int1}^{(1)} & \cdots & x_{int6}^{(1)} \\ \vdots & \ddots & \vdots \\ x_{int1}^{(63)} & \cdots & x_{int6}^{(63)} \end{bmatrix}, X_{seg(63\times6)} = \begin{bmatrix} x_{seg1}^{(1)} & \cdots & x_{seg6}^{(1)} \\ \vdots & \ddots & \vdots \\ x_{seg1}^{(63)} & \cdots & x_{seg6}^{(63)} \end{bmatrix}$$

where the variables x_{int1} to x_{int6} and x_{seg1} to x_{seg6} are the averages of the six driving behavioral measures for intersection- and segment-related events, respectively, and each matrix has 63 rows representing the 63 drivers. These two matrices (i.e., X_{int} and X_{seg}) were generated from a matrix (i.e., $X_{all\text{-}split}$) that combined all the intersection and segment training examples. This matrix was constructed to investigate the extent to which the drivers could be classified based on their proximity to intersections (i.e., location classification). This matrix (i.e., $X_{all\text{-}split}$) could be represented as shown below.

$$X_{all-split(126\times7)} = \begin{bmatrix} x_{int1}^{(1)} & \cdots & x_{int7}^{(1)} \\ \vdots & \ddots & \vdots \\ x_{int1}^{(63)} & \cdots & x_{int7}^{(63)} \\ x_{seg1}^{(1)} & \cdots & x_{seg7}^{(1)} \\ \vdots & \ddots & \vdots \\ x_{seg1}^{(63)} & \cdots & x_{seg7}^{(63)} \end{bmatrix}$$

The reason behind the above-mentioned location classification was to explore any potential variation in the driver behavior at intersections and on segments and to investigate to what extent this variation could be detected. Moreover, this classification was proposed to analyze the dissimilarity within each group (i.e., intersection and segment) on an aggregate level other than the driver individual level difference which was discussed in the previous chapter. Furthermore, this initial classification step could give some guidance on the suitability of the *K*-means for this type of classification. The *K*-means has been used extensively in the literature for drivers' classification as discussed in Chapter Two.

Generally, the proposed classification procedure, as an unsupervised learning procedure, consisted of two main steps; classification features extraction using PCA, and classification using *K*-means clustering.

PCA is a widely used method in several transportation applications for data reduction and feature extraction. For example, PCA was used to extract features from driver behavioral attributes that were collected on curves using a driving simulator (Chandrasiri et al., 2016, 2012). The extracted features were then used to classify drivers based on their driving skills (i.e., high, and low/average skilled drivers). Moreover, PCA was used to address multicollinearity and to extract the maximum information from a survey on drivers personalities (Guo and Fang, 2013). PCA was also used to represent most of the observed variance in a dataset into a smaller number of factors when using GPS data to investigate the behavior of commercial medium-sized truck drivers (Wu et al., 2016). Moreover, Wu et al. (2013) used PCA to identify the factors that influence the crash rates in urban expressways for different traffic conditions. The factors were considered to have a significant effect on crashes based on their weights in the principal components. Since each principal component has a linear combination of the studied factors, the factor with high weight (i.e., coefficient) in the top variance-principal components will have a significant effect on the crash occurrence (Wu et al., 2013). The PCA was utilized in the proposed classification procedure in this dissertation to extract the maximum amount of information while avoiding multicollinearity in the driving behavioral measures.

This chapter tackled the classification of the drivers based on their behavior and their location to shed light on the extent by which driving behavior could change due to their proximity to intersections. The PCA was used to extract a number of features (i.e., Principal Components) from the six indicators based on an eigenvalue of greater than 1.00. For more details about the PCA, readers are referred to (Jolliffe, 2002). Before performing the PCA, the data matrices (e.g., $X_{all\text{-}split}$) were normalized to a zero mean and unit variance. The mean normalization was carried out based on the following equation.

$$\bar{X}_i = \frac{X_i - \mu_i}{\sigma_i} \qquad (5)$$

where $\bar{X}_i$ is the normalized data matrix for driver i, and μ_i and σ_i are the mean and the standard deviation respectively of the data matrix X_i for each driver *i*. There are four data matrices available for this analysis (i.e., $X_{all\text{-}split}$, X_{all} , X_{seg} , and X_{int}) as discussed above.

The Principal Components (PCs) were estimated using MATLAB (The Mathworks Inc., 2016), and the main outputs of the analysis were the PCs for the training examples, the loadings (i.e., the coefficients of the six indicators) for each PC, and the eigenvalues for each PC.

A shown in Figure 25, the selected PCs were used to classify the drivers based on their location into intersection- and segment-related behavior and they were also classified into aggressive, normal, and cautious drivers using the *K*-means clustering (Hastie et al., 2009). The *K*-means clustering is a widely used technique and was used because of its flexibility in terms of selecting the number of the clusters. Moreover, the results of the *K*-means could be interpreted easily unlike the other clustering techniques such as Expectation Maximization. Also, the *K*-means is an unsupervised clustering technique unlike the nearest neighbor classification.

Each cluster resulting from the *K*-means had a group of points (each point represented a driver). This group of points was close to a specific centroid (i.e., cluster's centroid), which characterized the group and it was assumed that all the points within a cluster had homogenous attributes. The clustering analysis was performed for identifying drivers based on their proximity to intersections, and for driver behavior classification. Since the locations of the braking events were already known, the results of the clustering could be easily verified for the location classification, which will allow verifying the suitability of the procedure for the classification.

5.3 LOCATION CLASSIFICATION

To classify drivers based on their proximity to intersections, PCA was used to extract features (i.e., PCs) from the data matrix $X_{all\text{-}split}$. This matrix included the aggregated values of the six behavioral measures for each driver for each location separately (i.e., intersections and segments). As shown in Table 6, the first two PCs had eigenvalues greater than 1, and they accounted for more than 74% of the variance in the data. The loadings of these two PCs (i.e., PC1 and PC2) are shown in Table 7. It is worth mentioning that the relative speed and acceleration that were extracted for the incidents of interest, as explained earlier in Chapter Three, had negative signs. The negative sign of the relative speed means that the host vehicle was approaching (i.e., getting closer to) the leading vehicle and, for acceleration, means that the host vehicle was braking.

Table 6 *The results of the PCA of the data matrix $X_{all\text{-}split}$*

	All-split		
	Eigenvalues	**Percentage**	**Cumulative Percentage**
PC1	**3.148**	53.31%	53.31%
PC2	**1.258**	21.31%	**74.62%**
PC3	0.926	15.67%	90.29%
PC4	0.356	6.03%	96.32%
PC5	0.192	3.25%	99.57%
PC6	0.073	1.23%	100.81%

The *K*-means clustering was then used to classify the aggregated segment- and intersection-related events for the drivers using PC1 and PC2. The results are shown in Figure 26 where 126 points representing 63 drivers were plotted, and the clusters were identified. In this figure, each driver had two points one for segment- related events and another point for intersection-related events. The plot shows the segment- and intersection-related events for each driver and there are two distinct clusters with centroids of [-1.937, 0.078] and [1.317, -0.0529], respectively. Based on the grouping process of the events, the classification accuracy was estimated to be 90.5% since 12 points were misclassified as intersection-related events. The figure shows that the intersection-related events were clustered on the right-hand side of the plot with positive values of PC1 and the other event cluster (i.e., segment-related events) on the left-hand side of the plot.

Table 7 *The loadings for the first two Principal Components of the data matrix* $X_{all\text{-}split}$

	PC1	**PC2**
Following Distance	-0.485	0.305
Relative Speed	-0.477	-0.357
Acceleration	-0.435	0.322
Headway	0.225	0.727
TTC	-0.534	0.011
Jerk	-0.106	0.383

Based on the clustering results, the difference of PC2 values in clusters' centroids was minimal, and the clustering depended mainly on the values of PC1. Therefore, the loadings of PC1 were investigated to interpret the clustering results in terms of driver behavior. As shown in Table 7, the loadings of the first PC were all negative except the headway. Moreover, the highest magnitude of these loadings was for TTC followed by range, relative speed, acceleration, and jerk (which had a minimal effect on the PC1 value). Intuitively, the more the TTC value increased, the lower the

PC1 value would be. The same explanation for the TTC applies to the following distance. On the other hand, the larger the decrease in the relative speed, the lower the PC1 value and the same applies to acceleration. The interpretation of large TTC and following distance and small relative speed and acceleration values were that the driver had a safe maneuver and was less likely to be involved in a collision. Consequently, from a driver behavior perspective, when the PC1 value is negative; the driver had *less aggressive* behavior. Accordingly, it was concluded that the drivers are more likely to exhibit relatively aggressive behavior when approaching intersections when compared to driving on segments.

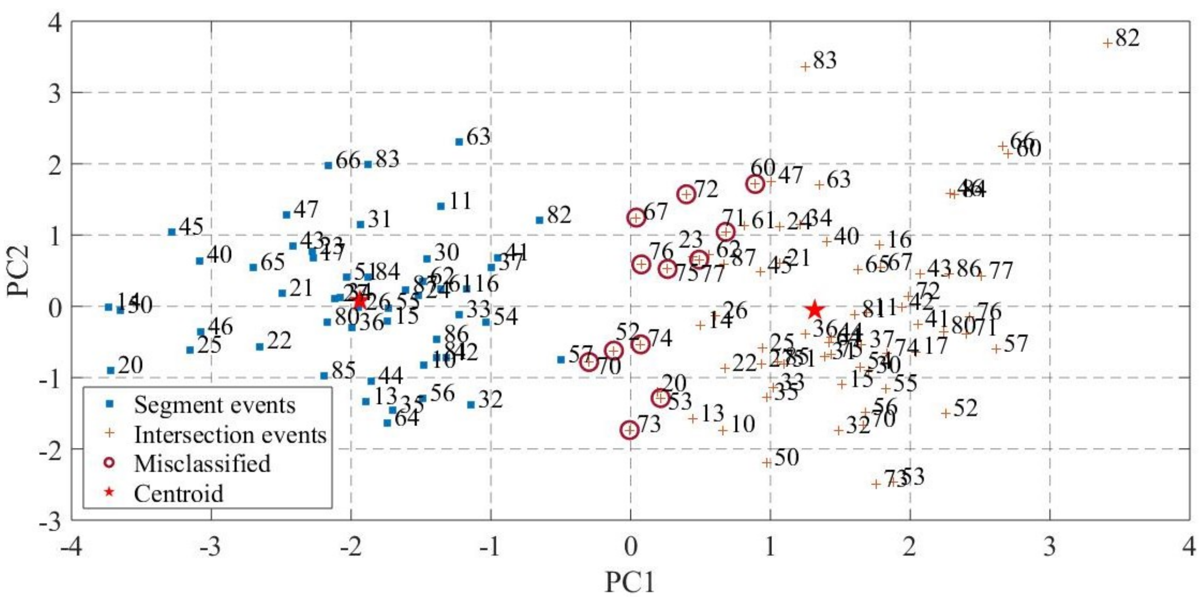


Figure 26. The *K*-means clustering results for segment- and intersection-related events.

Although the classification procedure misclassified some training examples as intersection events, the accuracy was still reasonable, which highlighted the suitability of the classification procedure. Moreover, the results showed in Figure 26 indicated that there was not only a tangible difference for the drivers in perceiving an intersection ahead but also this difference could be identified on an aggregate level.

Furthermore, the misclassified training examples shed light on the variation among each event group (i.e., intersection and segment groups). Therefore, the development of a context-aware DAS should consider the variation of the driver behavior due to the proximity to an intersection and the variation among the drivers. To capture the variation of the driver behavior due to the proximity to intersections, the infrastructure detection algorithm in Chapter Four was proposed to guarantee the consideration of the driver's needs while driving near intersections. For the variation among the drivers, the driver classification algorithm was proposed to classify the drivers into cautious, normal, and aggressive drivers to accommodate the driver variation in the context-aware DAS. Moreover, stakeholders could use such an algorithm to reward good driving while educating aggressive drivers or even incenting them to improve their driving behaviour.

5.4 BEHAVIORAL CLASSIFICATION

The behavior classification is the second algorithm (i.e., driver classification algorithm) in the context identification layer. This algorithm is concerned with classifying the driver into cautious, normal, or aggressive after detecting the influence of the infrastructure on the driver. Hence, the driver will be first checked whether he or she is influenced by an intersection (using the infrastructure detection algorithm) then the driver will be classified into cautious, normal, and aggressive. Therefore, for the driver classification algorithm, the classifications of the drivers at intersections and on segments are the only needed classifications. However, the behavioral classification, which will be discussed in this section, will classify the drivers three times as shown in Figure 25 to verify the classification results and to explore the dissimilarities between the classifications.

Moreover, to better interpret the misclassified events in the location classification, this behavioral classification was carried out separately for each event type (i.e., intersection-related and segment-related events). The PCs were extracted from the data matrices X_{all}, X_{seg}, and X_{int} for the classification. The results of the PCA, presented in Table 8, indicated that the first two PCs extracted from the three matrices had eigenvalues that exceeded 1 and accounted for at least 64% of the variance in the data. Therefore, PC1 and PC2 for the three cases were chosen for clustering using *K*-means. As shown in Table 9, the dominant variable loading for PC1 was the TTC in case of all events (case a) and segment-related events (case b). The TTC was followed by range, relative speed, and acceleration in (case a) and was followed by relative speed, acceleration, and range in (case b). All four variables (i.e., TTC, range, relative speed, and acceleration) had a negative correlation with PC1 for both cases. As explained before in the location classification, the large positive values of PC1 would be an indication of aggressive driving behavior for (case a) and (case b). Therefore, in Figure 27 and Figure 28, the drivers in group 1, 2, and 3 were classified, respectively, as cautious, normal, and aggressive drivers.

Moreover, in the form of contour maps (or heat maps), Figure 30 and Figure 31 show the relationship between the behavioral classification results for (case a) and (case b) and the highest loading of the first PC which was the TTC for both cases. As shown in these figures, the TTC values for group 1 (i.e., the cautious drivers) was larger than group 2 (i.e., the normal drivers). The same trend was noticed when comparing groups 2 and 3 where the TTC values of group 2 were larger than the TTC values of group 3 (i.e., aggressive drivers). For example, as shown in Figure 30, the centroid of group 3 in (case a) had a TTC value around 19 seconds, but the TTC value of group 2 was around 33 seconds. These values mean that the drivers in group 3 were relatively aggressive than the drivers in group 2. This insight supports the interpretation provided earlier in this section for the PCs loadings of both cases.

For intersection-related events (case c), the highest loading for PC1 was the headway followed by relative speed, range, and acceleration. All these loadings had a negative correlation with PC1 except relative speed which had a positive correlation with PC1. For PC2, the highest loading was TTC followed by acceleration, and relative speed and all of them had a negative correlation with PC2. Since PC1 accounted for the maximum variation in the data (i.e., 43.8%) as shown in Table 2, PC1 was used as the main component to interpret the results in Figure 29 (case c) and then PC2 values. For drivers in group 1, the results indicated cautious driving behavior because PC1 values were all negative. On the other hand, drivers in group 2 and 3 had positive PC1 values which means the drivers in group 2 and 3 were more aggressive than group 1 drivers. However, as shown in Figure 29, drivers in group 2 had negative PC2 values and drivers in group 3 had positive PC2 values. Again, as explained before, the negative PC2 values resulted from large values of TTC and range, and small values in the magnitude of acceleration and relative speed, which indicates less aggressive driving behavior. Thus, the drivers in group 2 who had negative PC2 values were less aggressive than the drivers in group 3 who had positive PC2 values. Hence, the drivers in group 2 and 3 were classified as normal and aggressive drivers, respectively. This interpretation aligns with the contour map results, which is presented in Figure 32. In this figure, the headway values of the centroids of groups 2 and 3 were about 1.55 and 1.35 seconds, respectively. Since the headway had the highest PC loading among the other behavioral measures in PC1, this result indicated that the drivers in group 3 were more aggressive than the drivers in group 2. Hence, groups 1, 2, and 3 were classified as cautious, normal, and aggressive drivers, respectively.

Table 8 *The results of the PCA of the data matrices X_{all}, X_{seg}, and X_{int}*

	All (a)			**Segments (b)**			**Intersections (c)**		
	Eigenvalues	**Percentage**	**Cumulative Percentage**	**Eigenvalues**	**Percentage**	**Cumulative Percentage**	**Eigenvalues**	**Percentage**	**Cumulative Percentage**
PC1	**2.421**	41.00%	41.00%	**2.541**	43.04%	43.04%	**2.586**	43.79%	43.79%
PC2	**1.512**	25.61%	**66.61%**	**1.260**	21.35%	**64.38%**	**1.505**	25.49%	**69.29%**
PC3	0.966	16.37%	82.98%	0.914	15.48%	79.86%	0.966	16.36%	85.64%
PC4	0.533	9.02%	92.00%	0.624	10.57%	90.43%	0.540	9.14%	94.78%
PC5	0.366	6.19%	98.19%	0.455	7.71%	98.15%	0.190	3.22%	98.01%
PC6	0.107	1.81%	100.00%	0.109	1.85%	100.00%	0.118	1.99%	100.00%

Table 9 *The loadings for the first two Principal Components of X_{all}, X_{seg}, and X_{int}*

	All (a)		Segments (b)		Intersections (c)	
	PC1	**PC2**	**PC1**	**PC2**	**PC1**	**PC2**
Range	-0.463	0.415	-0.386	0.556	-0.515	-0.299
Relative Speed	-0.452	-0.468	-0.484	-0.372	0.513	-0.373
Acceleration	-0.426	0.334	-0.423	0.285	-0.345	-0.440
Headway	0.167	0.650	0.254	0.636	-0.574	0.062
TTC	-0.598	0.033	-0.568	0.105	0.077	-0.740
Jerk	0.122	0.272	0.223	0.236	-0.134	0.164

The results of the clustering revealed that 12 drivers, who were misclassified in the location classification earlier, were clustered in the same group (i.e., group 3) for (case a) and (case b) as shown in *Figure 27* and *Figure 28*. Upon closer examination, the attribute of these clusters indicates that this group of drivers (i.e., group 3 in *Figure 27* and *Figure 28*) was more aggressive than the rest. These findings highlight that the segment-related events were misclassified as intersection-related because of the drivers' relatively aggressive behavior.

Furthermore, Table 10 shows the clustering results for all the cases and the number and the percentages of the drivers in each cluster. As shown in *Figure 27*, *Figure 28*, and *Figure 29*, and in Table 10 *Clustering results of drivers classification*.

The classifications in (case a) and (case b) were similar when compared to the intersection-related events (case c). For instance, the thirteen drivers who were identified as aggressive drivers in (case a) were the same aggressive drivers in (case b) as discussed earlier. Moreover, some drivers were identified as cautious drivers (10 drivers) and normal drivers (31 drivers) in both (case a) and (case b). Table 10 *Clustering results of drivers classification* shows the drivers' distribution for (case a) and (case b) which is consistent with the proposed distribution by Jensen et al. (2011) where the normal drivers represent 50% and cautious and aggressive drivers represent 25%.

In contrast, there was a tangible difference between the classification in these two cases (i.e., case a and case b) and (case c) where more aggressive drivers were identified. As shown in *Figure 29* and in Table 10, the same 10 drivers were identified as aggressive drivers in all the cases. However, 15 more drivers (i.e., 25 drivers in total) were identified as aggressive drivers near intersections. Those 15 drivers were identified as either normal drivers (12 drivers) or cautious drivers (3 drivers) in (case a) and (case b). This is an indication that more drivers tend to be more aggressive at intersections when compared to their driving behavior elsewhere on the road network.

Table 10 *Clustering results of drivers classification*

	All (a)		**Segment (b)**		**Intersection (c)**	
	Centroid	No. of drivers	Centroid	No. of drivers	Centroid	No. of drivers
Group 1	(-1.998,-0.005)	14 (**22.2%**)	(-1.903,0.608)	15 (**23.8%**)	(-2.376,0.224)	13 (**20.6%**)
Group 2	(-0.065,0.077)	36 (**57.2%**)	(-0.092,-0.190)	35 (**55.6%**)	(0.203,-0.956)	25 (**39.7%**)
Group 3	(2.331,-0.208)	13 (**20.6%**)	(2.443,-0.191)	13 (**20.6%**)	(1.033,0.840)	25 (**39.7%**)

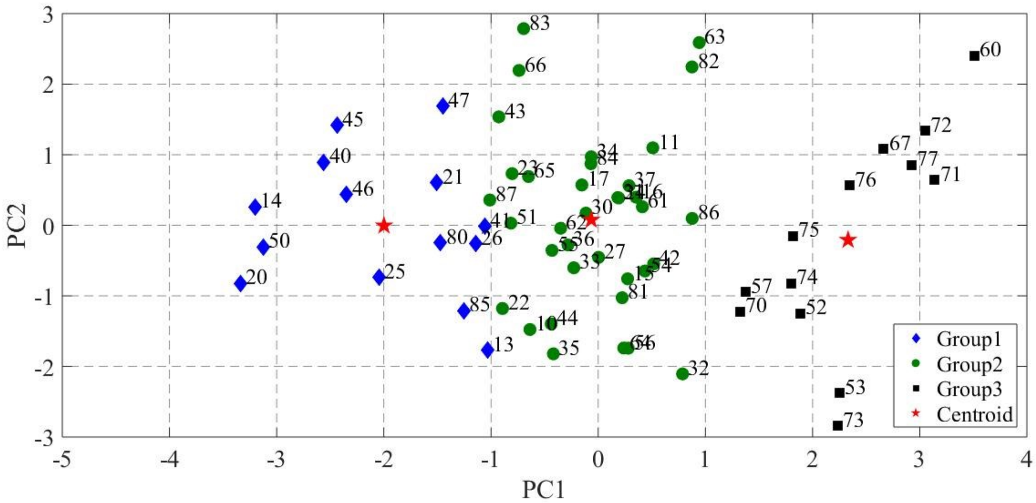


Figure 27. Behavior classification results with the drivers' IDs for case a.

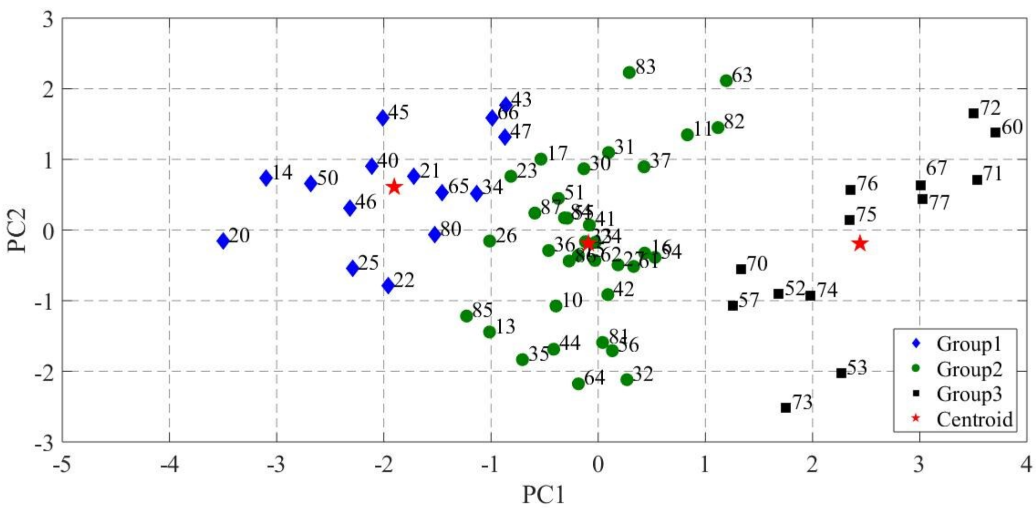


Figure 28. Behavior classification results with drivers' IDs for case b.

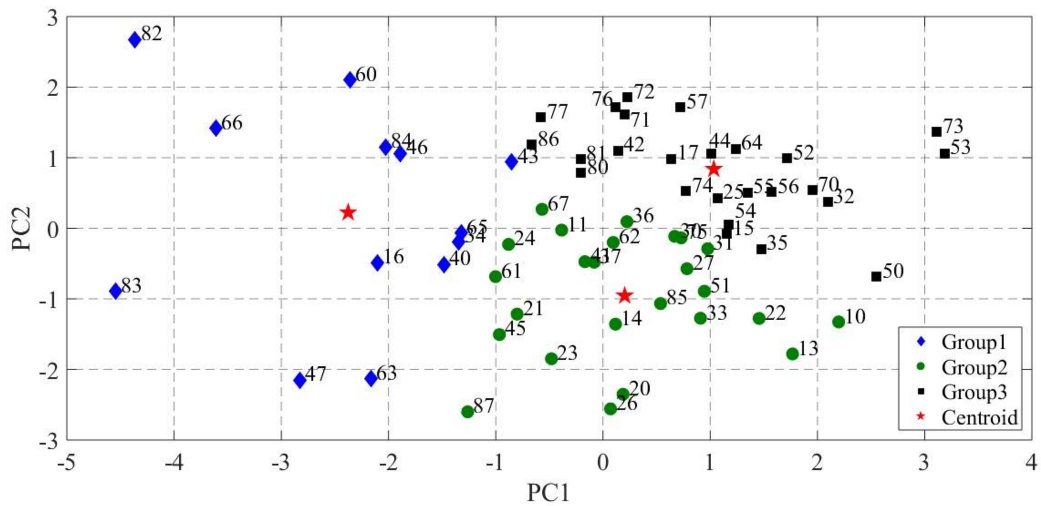


Figure 29. Behavior classification results with drivers' IDs for case c.

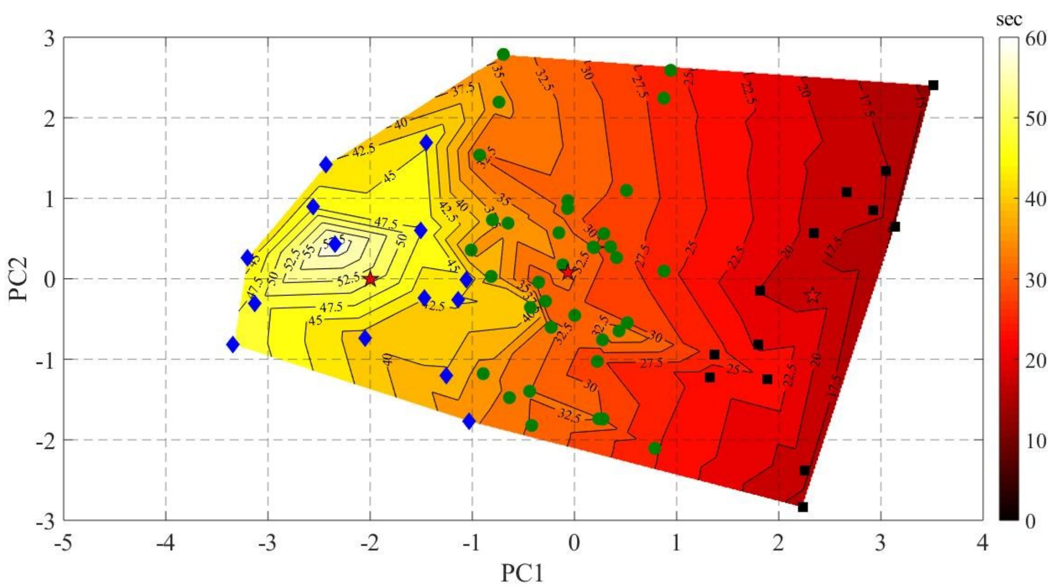


Figure 30. Contour map for the TTC values for case a.

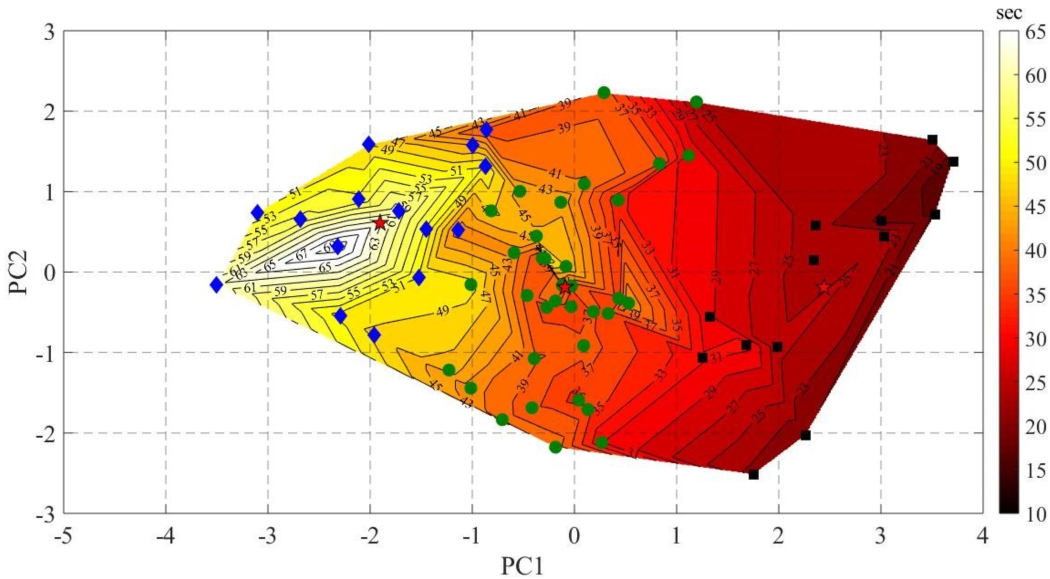


Figure 31. Contour map for the TTC values for case b.

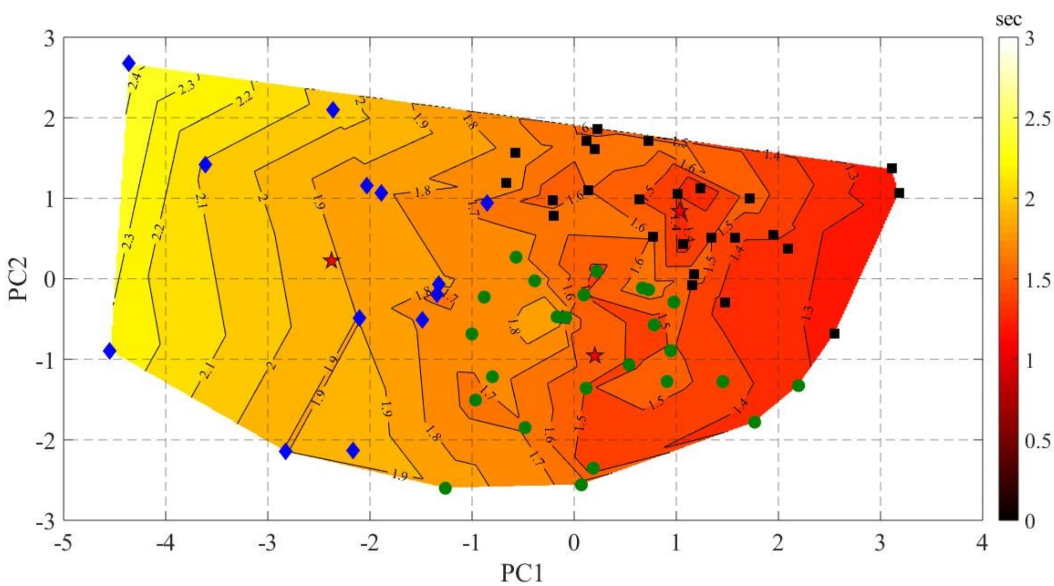


Figure 32. Contour map for the headway values for case c.

5.5 SUMMARY AND DISCUSSION

This chapter focused on the development of the driver classification algorithm, which was proposed as the second algorithm in the context identification layer of the reasoning subsystem in a context-aware DAS. The primary purpose of this algorithm was to classify the drivers into cautious, normal, and aggressive in order to accommodate the variation in driver behavior into the context-aware DAS. Based on the output of the infrastructure detection algorithm, as discussed earlier, the driver classification algorithm could classify the drivers into three classes (i.e., cautious, normal, and aggressive) based on their location (i.e., at intersections or on segment) as shown in Figure 17. Then, the core algorithm of the DAS could provide the warning or the service to the drivers based on their class, which will enhance the adaptability of the context-aware DAS. To develop the driver behavior classification algorithm, a classification procedure, which includes two main steps; the PCA, and *K*-means clustering was proposed. The classification was performed at different levels (i.e., location and driver behavior classification) to verify the results and to explore the dissimilarities between the drivers' groups. However, the driver classification algorithm was based on the behavioral classification only. For the location grouping, the classification accuracy reached about 90% based on the aggregate data from the sixty-three drivers. The results indicated that there was a tangible difference among the drivers in perceiving an intersection ahead. Nonetheless, the interesting finding in this classification (i.e., location classification) was the segment-related events, which were misclassified as intersection-related events. These misclassified events did not only highlight the variation among each group of drivers but also raised a question about the reasons behind this misclassification. The answer to this question was addressed by the behavioral classification, which was carried out on different levels as shown in Figure 24 and Figure 25.

The behavioral classification showed that all the misclassified events belonged to the group of aggressive drivers. Also, the behavior classification results indicated that TTC, range, relative speed, and acceleration were the dominant measures influencing the classification of all drivers and drivers on segment-related events. For intersection-related events classification, headway, range, acceleration, and TTC had more influence in identifying aggressive drivers at the proximity of intersections. Moreover, there was an agreement between drivers' classification for all drivers' events and drivers' segment events. On the contrary, drivers' intersection events classification was dramatically different from the above-mentioned cases since more drivers tended to be more aggressive. These insights highlighted the importance of considering the location of the drivers when classifying them, which was not considered in the previous studies such as (Osafune et al., 2017; Rong et al., 2011; Wang et al., 2010a). Therefore, the context identification layer with its components is crucial for the design context-aware DAS since this layer will acknowledge the changes of driving context in terms of surrounding infrastructure and how the drivers could perceive such changes.

6. SUMMARY

This chapter summarizes the main conclusions, contributions, and future research and limitation of this dissertation.

6.1 CONCLUSIONS

The results of this dissertation revealed some crucial insights into the driver car-following behavior. These insights could be summarized in the following points.

- The car-following events during braking could be quantified to provide an overview of the driver behavior at different locations on the road network.
- The definition of the intersection influence area based on the SSD at different speed groups (i.e., low, medium, and high-speeds) was efficient and could separate the intersection-related behavior from the segment-related behavior.
- There was a significant difference between the driver behavior at intersections and on segments based on the behavioral measures that were examined in this dissertation. These behavioral measures were following distance, relative speed, acceleration, TTC, headway, and jerk.
- The drivers, in general, tend to be relatively more aggressive at intersections when compared to their behavior on segments. This aggressiveness was identified with respect to the segment behavior as a baseline condition and was defined based on the above-mentioned behavioral measures.
- The driving behavior near the intersections could be distinguished from the driving behavior on segments for **each individual driver** using deep learning (i.e., deep neural network) with enough accuracy.

- At least three of the four variables (i.e., the following distance, the headway, the acceleration, and the relative speed) should be included as the input variables to obtain high intersection behavior detection accuracy. On other hand, TTC and jerk were not among these variables which means that they were less impacted with the proximity to intersections. However, TTC had a significant impact on the classification of drivers as discussed before.
- The acceleration had the highest impact on the detection accuracy which could be translated as the acceleration was the highest of the behavioral measures to be influenced by driving in the proximity to intersections.
- The intersection-related behavior could be easily separated from the segment-related behavior on **an aggregate level** and after combining all the behavioral measures using the PCA. This insight emphasized the substantial difference between the driver behavior at different locations on the road network.
- There was an agreement between the driver classification in general and the driver classification using segment-related events only. However, there was a significant difference between these two cases and the driver classification using intersection-related events. This indicated that even if the general driver classification (i.e., regardless the driver location) was similar to the classification on segments, this general classification will not be applicable at intersections. In other words, the general classification should not be used on the zones which have high intersection density such as downtown cores.
- Some drivers showed a normal (i.e., moderate) driving behavior on segments but had a relatively aggressive behavior at intersections. This insight highlighted the importance of the location-based driver classification.

- Although the TTC did not have a tangible effect on the detection of intersection-related behavior, the TTC had a considerable effect on the driver classification at different levels (i.e., on segments, for the general classification, and for the location classification).

These insights were integrated into the architecture of the context-aware DAS by adding a context identification layer to the reasoning subsystem. The context identification layer was proposed to detect the impact of the surrounding infrastructure change (i.e., from driving on segments to driving near intersections) on the driver and to classify the drivers based on their location on the road network. This layer will enhance the adaptability of the entire context-aware DAS to the driving context including the driver behavior variation, the change in the surrounding environment, and the interaction with the surrounding vehicle (described as car-following interaction).

The proposed layer encompasses two main algorithms, namely, the infrastructure detection algorithm and the driver classification algorithm. These two layers were proposed to work in sequence where the driver classification algorithm (i.e., the second algorithm) depended on the results of the infrastructure detection algorithm (i.e., the first algorithm). Nevertheless, the context identification algorithm still could have one of these algorithms only. For instance, the infrastructure detection algorithm could work alone in CV environment to detect when is the best time to send the driver a warning message about an intersection. The best time for this warning could be defined as when the algorithm first detects that the driver is influenced by the intersection. Therefore, this algorithm could supplement the CV applications even in a V2I environment. For example, when a driver is approaching an intersection in a V2I communication environment, a typical warning message might be sent to the driver once the vehicle is connected to the intersection's infrastructure. However, this typical message could be a late/early warning because the driver could be influenced by the intersection before/after the vehicle gets connected.

Therefore, the development of context-aware systems which detect the effect of the surrounding environment on the driver is a crucial component for reliable DAS. In future where fully AV are used in the roads, the driver behavior insights presented in this dissertation could be used for improving the riding comfort of AV passengers.

The same applies to the second algorithm (i.e., the driver classification algorithm) which could operate alone in the context identification layer using driving data from different environments (e.g., CV or AV). This algorithm could customize the warning message based on the drivers class. Moreover, this algorithm alone could be the core of car insurance applications which could offer the normal drivers some incentives (e.g., low insurance premiums) for their "good" driving behavior. Furthermore, this algorithm could help the stakeholders to educate the drivers about their behavior such as giving more driving tips to the aggressive drivers who might not know that they are driving in an aggressive behavior. Also, the algorithm could help the stakeholders to identify clear cut-offs for aggressive driving which could supplement the traffic enforcement programs.

One of the main advantages of the proposed layer is that it could be customized based on each individual driver behavior (i.e., driver-based). In other words, this layer will learn from the driver behavioral measures (i.e., data-driven) to differentiate between intersection- and segment-related behaviors. These two properties, which are driver-based and data-driven, will allow the transferability of the DAS to other regions than the region where the NDS data was collected.

6.2 CONTRIBUTIONS

This dissertation presented several contributions to state of the art in both the driver behavior study and in the development of the context-aware DAS. The significant contributions of this dissertation are highlighted in the following points.

- A clear definition of the intersection influence area based on measurable design parameters (i.e., SSD and the speed limit) was presented. This dissertation defined the intersection influence area as the area where the driver behavior is affected substantially by an intersection ahead which will impact the driver's maneuvering decisions including car-following, braking, and lane changing decisions. This influence area includes the physical area of the intersection which is bounded by the stop lines and the area upstream the stop line of the intersection limited by the SSD which is estimated based on the speed of the driver and the speed limit.
- The probability density functions with the name and the parameters for each distribution of the behavioral measures were listed for both intersection- and segment-related events. These behavioral measures included the following distance, relative speed, and headway between the host vehicle and the leading vehicle, TTC, acceleration, and jerk. These density functions could be used for driver behavior modeling and for calibrating and validating microsimulation models as suggested in (Tawfeek et al., 2018).
- A context identification layer was proposed as an addition to the reasoning subsystem of the context-aware DAS to enhance the adaptability of the system. The first component of this layer (i.e., infrastructure detection algorithm) provided a new aspect to the context-aware DAS which was the detection of the change in driver behavior based on the microscopic alteration in the driving context. This microscopic alteration in the driving context included the change in the surrounding environment (i.e., described as the proximity to intersections).
- The second component of the context identification layer (i.e., driver detection algorithm) was proposed to accommodate the location-based variation in the driver behavior into the

context-aware DAS. Moreover, this algorithm combined six behavioral measures which were extracted from the NDS for the driver classification, unlike the previous studies which focused on using two behavioral measures from the NDS data.

- The driver classification results highlighted that the current driver classification procedures should not be generalized on all the locations of the road network such as the intersections since the results indicated that more drivers were considered as aggressive at intersections while they were identified as normal drivers in the general classification. This insight could contribute to a more reliable car insurance application which should differentiate between driving within the cities' downtowns where the density of the intersections is high and driving on the highways where the density of the intersections is significantly lower than the downtowns.
- This dissertation shed light on the importance of integrating the NDS data with the road network maps to deeply investigate the driver behavior. The pairing of the NDS and the road network using ArcGIS, for instance, is a very promising area of research which will allow insightful findings in driver behavior understanding.

6.3 FUTURE RESEARCH AND LIMITATIONS

Despite the extensive research carried out in this dissertation, driver behavior research still needs more analysis and investigation due to the complexity of human behavior in general, and human interactions within the driving environment in particular.

This dissertation provides different probability density functions for several behavioral measures for both intersections and segments. These functions could be efficiently utilized for microsimulation model calibration and validation. Moreover, these functions could be considered for the development of car-following models or for enhancing the current models. Furthermore, the proposed definition of the intersection influence area could be used for the analyses of intersection and intersection-related crashes and could be considered for the investigation of dilemma zones of signalized intersections. However, one limitation of the analyses is the intersections were not split into signalized and unsignalized intersections, which was left for the future research. Also, for the classification procedure, the segment-related events could be split into curves and straight segments or weaving and non-weaving sections for more in-depth insights into driver behavior.

The NDS data available for this dissertation did not include any information about driver demographics, such as the age and gender, secondary driving tasks, nor did it include weather information. The inclusion of such information will provide interesting additional insights. In addition, the impact of different traffic conditions in terms of traffic flow should be explored. Using the data of 63 drivers, which is a comparatively large sample of drivers to the previous research in this area, the proposed context identification layer could be trained using more data for better accuracy. For the infrastructure detection algorithm, the test accuracy of the trained neural networks was not as high as the training accuracy because a small number of training examples were extracted for some drivers, which affected the average total accuracy that was estimated for all the drivers. Moreover, the neural networks were trained based on the available behavioral measures. Therefore, future research can consider more behavioral measures for the training process, such as the throttle opening or the throttle pressure, the brake pedal pressure, Time Exposed Time-to-Collision (TET), Time Integrated Time-to-Collison (TIT), etc.

On the other hand, future research could also investigate the transferability of the proposed context identification layer to other types of drivers, such as taxi drivers, bus drivers, and heavy vehicle drivers. This future research direction will provide stakeholders with means of drivers' education to enhance the service provided by those drivers. Moreover, the results of this dissertation could be extended to provide some guidelines and thresholds not only for enforcement but also to coach drivers who have aggressive driving behavior. However, this extension needs more data from more drivers to present clear and reliable enforcement and coaching programs. Furthermore, the findings of this dissertation could provide significant enhancement to the current car insurance programs, which offer lower premiums to the "good" drivers who are not identified as aggressive. The enhancement will mainly benefit users who drive their cars within the downtown cores or within high intersection density areas.

Various machine learning techniques were used to investigate the driver behavior in this dissertation. Although the techniques used (i.e., neural networks, PCA, and *K*-means) in the literature are extensively adapted to study different aspects of driver behavior, it is worth exploring other machine learning techniques and comparing the results. Also, the training of the neural network could account for regularization to avoid any possible overfitting to the training data, and hence, enhance the testing accuracy. Moreover, the driver classification algorithm was based on an unsupervised machine learning technique, which means that the results were not verified with ground truth cases due to the limitation of the data as discussed earlier. However, the promising results obtained from the classification suggested that the proposed procedure is valid. Therefore, it would be interesting to compare the results of the proposed classification procedure to the history of the drivers and their experiences.

Finally, this dissertation sheds light on the importance of integrating NDS with the road network map to deeply further investigate driver behavior. The pairing of the NDS and the road network, using ArcGIS, LiDAR data, or other means, is an up-and-coming area of research that will allow insightful findings towards understanding driver behavior.

REFERENCES

AASHTO, 2011. A Policy on Geometric Design of Highways and Streets 942. doi:10.1029/2004JC002361

AASHTO, 2010. Highway safety manual, first edition, with 2014 supplement.

Abdelwahab, H., Abdel-Aty, M., 2002. Artificial neural networks and logit models for traffic safety analysis of toll plazas. Transportation Research Record 1784 , 115–125. doi:10.3141/1784-15

Abe, G., Richardson, J., 2005. The influence of alarm timing on braking response and driver trust in low speed driving. Safety Science 43 9 , 639–654. doi:10.1016/j.ssci.2005.04.006

Adams, L., 1994. Review of the literature on obstacle avoidance maneuvers: Braking versus steering August , 22. doi:10.1186/1757-1626-2-9342.7.

Aghabayk, K., Sarvi, M., Young, W., 2012. Understanding the Dynamics of Heavy Vehicle Interactions in Car-Following. Journal of Transportation Engineering 138 December , 120820061029002. doi:10.1061/(ASCE)TE.1943-5436.0000463

Ahmed, M.M., Ghasemzadeh, A., 2018. The impacts of heavy rain on speed and headway Behaviors: An investigation using the SHRP2 naturalistic driving study data. Transportation Research Part C: Emerging Technologies 91, 371–384. doi:10.1016/j.trc.2018.04.012

Al-Sultan, S., Al-Bayatti, A.H., Zedan, H., 2013. Context-aware driver behavior detection system in intelligent transportation systems. IEEE Transactions on Vehicular Technology 62 9 , 4264–4275. doi:10.1109/TVT.2013.2263400

Alden, A.S., Mayer, B., Mcgowen, P., Sherony, R., Takahashi, H., 2016. Animal-Vehicle Encounter Naturalistic Driving Data Collection and Photogrammetric Analysis. SAE Technical Paper 2016.

Alhammad, A., Siewe, F., Al-Bayatti, A.H., 2012. An InfoStation-based context-aware on-street parking system. 2012 International Conference on Computer Systems and Industrial Informatics, ICCSII 2012. doi:10.1109/ICCSII.2012.6454358

Arbabzadeh, N., Jafari, M., 2018. A Data-Driven Approach for Driving Safety Risk Prediction Using Driver Behavior and Roadway Information Data. IEEE Transactions on Intelligent Transportation Systems 19 2 , 446–460. doi:10.1109/TITS.2017.2700869

Bagdadi, O., 2013. Assessing safety critical braking events in naturalistic driving studies. Transportation Research Part F: Traffic Psychology and Behaviour 16, 117–126. doi:10.1016/j.trf.2012.08.006

Bagdadi, O., Várhelyi, A., 2013. Development of a method for detecting jerks in safety critical events. Accident Analysis and Prevention 50, 83–91. doi:10.1016/j.aap.2012.03.032

Barber, P., Clarke, N., 1998. Advanced collision warning systems. IEE Colloquium 234 2 , 1–9.

Bärgman, J., van Nes, N., Christoph, M., Jansen, R., Heijne, V., Carsten, O., Dotzauer, M., Utech, F., Svanberg, E., Cocron, M.P., Forcolin, F., Kovaceva, J., Guyonvarch, L., Hibberd, D., Lotan, T., Winkelbauer, M., Sagberg, F., Stemmier, E., Gellerman, H., Val, C., Quintero, K., Tattegrain, H., Donabauer, M., Pommer, A., Neumann, I., Albert, G., Welsh, R., Fox, C., 2017. UDrive deliverable D41.1: The UDrive dataset and key analysis results. doi:[to be added]

Bejani, M.M., Ghatee, M., 2018. A context aware system for driving style evaluation by an ensemble learning on smartphone sensors data. Transportation Research Part C: Emerging Technologies 89, 303–320. doi:10.1016/j.trc.2018.02.009

Bella, F., Calvi, A., D'Amico, F., 2014. An empirical study on traffic safety indicators for the analysis of car-following conditions. Advances in Transportation Studies 1 SPECIAL ISSUE

, 5–16. doi:10.4399/97888548735442

Bella, F., Russo, R., 2011. A Collision Warning System for rear-end collision : a driving simulator study. Procedia - Social and Behavioral Sciences 20, 676–686. doi:10.1016/j.sbspro.2011.08.075

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., Winner, H., 2014. Three Decades of Driver Assistance Systems: Review and Future Perspectives. Intelligent Transportation Systems Magazine, IEEE 6 4 , 6–22. doi:10.1109/MITS.2014.2336271

Berndt, H., Wender, S., Dietmayer, K., 2007. Driver braking behavior during intersection approaches and implications for warning strategies for driver assistant systems. Proc. IEEE Intelligent Vehicles Symposium 245–251. doi:10.1109/IVS.2007.4290122

Bliss, J.P., Acton, S.A., 2003. Alarm mistrust in automobiles: How collision alarm reliability affects driving. Applied Ergonomics 34 6 , 499–509. doi:10.1016/j.apergo.2003.07.003

Boda, C.-N., Dozza, M., Bohman, K., Thalya, P., Larsson, A., Lubbe, N., 2018. Modelling how drivers respond to a bicyclist crossing their path at an intersection: How do test track and driving simulator compare? Accident Analysis & Prevention 111, 238–250. doi:10.1016/j.aap.2017.11.032

Böhmländer, D., Dirndorfer, T., Al-Bayatti, A.H., Brandmeier, T., 2017. Context-aware system for pre-triggering irreversible vehicle safety actuators. Accident Analysis and Prevention 103, 72–84. doi:10.1016/j.aap.2017.02.015

Boyraz, P., Hansen, J.H.L., 2009. Active vehicle safety system design based on driver characteristics and behaviour. International Journal of Vehicle Safety 4 4 , 330. doi:10.1504/IJVS.2009.032760

Canada Naturalistic Driving Study [WWW Document], 2018. URL http://www.canada-nds.net/

Carney, C., Harland, K.K., McGehee, D. V., 2016. Using event-triggered naturalistic data to examine the prevalence of teen driver distractions in rear-end crashes. Journal of Safety Research 57, 47–52. doi:10.1016/j.jsr.2016.03.010

Castignani, G., Derrmann, T., Frank, R., Engel, T., 2015. Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. IEEE Intelligent Transportation Systems Magazine 7 1 , 91–102. doi:10.1109/MITS.2014.2328673

Casucci, M., Marchitto, M., Cacciabue, P.C., 2010. A numerical tool for reproducing driver behaviour: Experiments and predictive simulations. Applied Ergonomics 41 2 , 198–210. doi:10.1016/j.apergo.2009.01.008

Chandrasiri, N.P., Nawa, K., Ishii, A., 2016. Driving skill classification in curve driving scenes using machine learning. Journal of Modern Transportation 24 3 , 196–206. doi:10.1007/s40534-016-0098-2

Chandrasiri, N.P., Nawa, K., Ishii, A., Li, S., Yamabe, S., Hirasawa, T., Sato, Y., Suda, Y., Matsumura, T., Taguchi, K., 2012. Driving skill analysis using machine learning: The full curve and curve segmented cases. 2012 12th International Conference on ITS Telecommunications, ITST 2012 542–547. doi:10.1109/ITST.2012.6425238

Chang, J., Yao, W., Li, X., 2017. The design of a context-aware service system in intelligent transportation system. International Journal of Distributed Sensor Networks 13 10 , 1–18. doi:10.1177/1550147717738165

Chang, L.Y., 2005. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. Safety Science 43 8 , 541–557. doi:10.1016/j.ssci.2005.04.004

Charalambous, C., 1992. Conjugate gradient algorithm for efficient training of artificial neural networks. Circuits, Devices and Systems, IEEE Proceedings G 139 3 , 301–310.

Chatterjee, I., Davis, G., 2014. Use of Naturalistic Driving Data to Characterize Driver Behavior in Freeway Shockwaves. Transportation Research Record: Journal of the Transportation Research Board 2434, 9–17. doi:10.3141/2434-02

Chen, R., Kusano, K.D., Gabler, H.C., 2015. Driver Behavior During Overtaking Maneuvers from the 100-Car Naturalistic Driving Study. Traffic Injury Prevention 16 May , S176–S181. doi:10.1080/15389588.2015.1057281

Chen, R., Sherony, R., Gabler, H.C., 2016. Comparison of Time to Collision and Enhanced Time to Collision at Brake Application during Normal Driving. SAE Technical Paper 2016-01-1448 . doi:10.4271/2016-01-1448

Chen, X., Chi, R., Wang, J., Pang, C., Xiao, Q., 2012. Driver safe speed model based on BP neural network for rural curved roads. IFIP Advances in Information and Communication Technology 370 AICT PART 3 , 92–102. doi:10.1007/978-3-642-27275-2_10

Chen, Y., Quddus, M., Wang, X., 2018. Impact of combined alignments on lane departure: A simulator study for mountainous freeways. Transportation Research Part C: Emerging Technologies 86, 346–359.

Chi, R.. b, Wang, J.., Yang, F.., Zhang, L.., 2010. A forward collision warning algorithm based on driver characteristics. ICCTP 2010: Integrated Transportation Systems: Green, Intelligent, Reliable - Proceedings of the 10th International Conference of Chinese Transportation Professionals 382, 640–652. doi:10.1061/41127(382)68

Chong, L., Abbas, M.M., Medina Flintsch, A., Higgs, B., 2013. A rule-based neural network approach to model driver naturalistic behavior in traffic. Transportation Research Part C: Emerging Technologies 32, 207–223. doi:10.1016/j.trc.2012.09.011

City of Ann Arbor, 2013. City of Ann Arbor Speed Limits Map [WWW Document]. URL http://www.a2gov.org/departments/engineering/Pages/default.aspx (accessed 4.12.17).

Cobb, S.V.G., Nichols, S., Ramsey, A., Wilson, J.R., 1999. Virtual reality-induced symptoms and effects (VRISE). Presence 8 2 , 169–186. doi:10.1086/250095

Colombaroni, C., Fusco, G., 2014. Artificial neural network models for car following: Experimental analysis and calibration issues. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 18 1 . doi:10.1080/15472450.2013.801717

Costa, M., Ferreira, B.Q., Marques, M., 2017. A Context Aware and Video-Based Risk Descriptor for Cyclists 1459–1464.

Cottrell, W.D., Mu, S., 2005. Utah Intersection Safety: Recurrent Crash Sites: Identification, Issues and Factors. University of Utah, Salt Lake City.

D'Orazio, T., Leo, M., Guaragnella, C., Distante, A., 2007. A visual approach for driver inattention detection. Pattern Recognition 40 8 , 2341–2355. doi:10.1016/j.patcog.2007.01.018

De Winter, J.C.F., van Leeuwen, P.M., Happee, R., 2012. Advantages and Disadvantages of Driving Simulators: A Discussion, in: Measuring Behavior. (Utrecht, The Netherland, pp. 47–50.

Debnath, A.K., Blackman, R., Haworth, N., Adinegoro, Y., 2017. Influence of Remotely Operated Stop–Slow Controls on Driver Behavior in Work Zones. Transportation Research Record: Journal of the Transportation Research Board pp 19–25. doi:10.3141/2615-03

Dingus, T.A., Hankey, J.M., Antin, J.F., Lee, S.E., Eichelberger, L., Stulce, K.E., McGraw, D., Perez, M., Stowe, L., 2015. Naturalistic Driving Study: Technical Coordination and Quality Control 370. doi:S2-S06-RW-1

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., Knipling, R.R.,

2006. The 100-Car Naturalistic Driving Study Phase II – Results of the 100-Car Field Experiment. Chart April , 442. doi:DOT HS 810 593

Doi, A., Butsuen, T., Niibe, T., Takagi, T., Yamamoto, Y., Seni, H., 1994. Development of a rear-end collision avoidance system with automatic brake control. JSAE Review 15 4 , 335–340. doi:10.1016/0389-4304(94)90216-X

Donmez, B., Boyle, L., Lee, J., 2007. SAfety VEhicles using adaptive Interface Technology (Task 4) Final Report: Phase 2 Distraction Mitigation Evaluation. Iowa City, IA: University of Iowa.

Dougherty, M.S., Kirby, H.R., Boyle, R.D., 1993. The use of neural networks to recognise and predict traffic congestion. Traffic engineering & control 34 6 , 311–4.

Eftekhari, H.R., Ghatee, M., 2018. Hybrid of discrete wavelet transform and adaptive neuro fuzzy inference system for overall driving behavior recognition. Transportation Research Part F: Traffic Psychology and Behaviour 58, 782–796. doi:10.1016/j.trf.2018.06.044

Evans, L., 1991. Traffic Safety and the driver, New York: Van Nostrand Reinhold . doi:10.2307/2076302

Fang, S.H., Fei, Y.X., Xu, Z., Tsao, Y., 2017. Learning Transportation Modes from Smartphone Sensors Based on Deep Neural Network. IEEE Sensors Journal 17 18 , 6111–6118. doi:10.1109/JSEN.2017.2737825

Farah, H., Koutsopoulos, H.N., Saifuzzaman, M., Kölbl, R., Fuchs, S., Bankosegger, D., 2012. Evaluation of the effect of cooperative infrastructure-to-vehicle systems on driver behavior. Transportation Research Part C: Emerging Technologies 21 1 , 42–56. doi:10.1016/j.trc.2011.08.006

Feng, F., Bao, S., Sayer, J.R., Flannagan, C., Manser, M., Wunderlich, R., 2017. Can vehicle longitudinal jerk be used to identify aggressive drivers? An examination using naturalistic driving data. Accident Analysis and Prevention 104, 125–136. doi:10.1016/j.aap.2017.04.012

FHWA, U.S.D. of T., 2009. Manual on Uniform Traffic Control Devices.

Fitzpatrick, C.D., Samuel, S., Knodler, M.A., 2017. The use of a driving simulator to determine how time pressures impact driver aggressiveness. Accident Analysis and Prevention 108, 131–138. doi:10.1016/j.aap.2017.08.017

Fitzpatrick, C.D., Samuel, S., Knodler, M.A., 2016. Evaluating the effect of vegetation and clear zone width on driver behavior using a driving simulator. Transportation Research Part F: Traffic Psychology and Behaviour 42, 80–89. doi:10.1016/j.trf.2016.07.002

Fitzpatrick, K., Carlson, P.J., Wooldrigge, M.D., Brewer, M.A., 2000. Design factors that affect driver speed on suburban arterials 164.

Fugiglando, U., Massaro, E., Santi, P., Milardo, S., Abida, K., Stahlmann, R., Netter, F., Ratti, C., 2018. Driving Behavior Analysis through CAN Bus Data in an Uncontrolled Environment. IEEE Transactions on Intelligent Transportation Systems. doi:10.1109/TITS.2018.2836308

Fujita, Y., Akuzawa, K., Sato, M., 1995. Radar brake system. JSAE Review 16(2), 95–101.

Gettman, D., Pu, L., Sayed, T., Shelby, S.G., 2008. Surrogate Safety Assessment Model and Validation : Final Report June , FHWA-HRT-08-051.

Ghasemzadeh, A., Ahmed, M.M., 2017. Drivers’ Lane-Keeping Ability in Heavy Rain: Preliminary Investigation Using SHRP 2 Naturalistic Driving Study Data. Transportation Research Record: Journal of the Transportation Research Board 2663 1 , 99–108. doi:10.3141/2663-13

Gilman, E., Keskinarkaus, A., Tamminen, S., Pirttikangas, S., Röning, J., Riekki, J., 2015. Personalised assistance for fuel-efficient driving. Transportation Research Part C: Emerging

Technologies 58 PD , 681–705. doi:10.1016/j.trc.2015.02.007

Guo, F., Fang, Y., 2013. Individual driver risk assessment using naturalistic driving data. Accident Analysis and Prevention 61, 3–9. doi:10.1016/j.aap.2012.06.014

Guo, F., Fang, Y., Antin, J.F., 2015. Older driver fitness-to-drive evaluation using naturalistic driving data. Journal of Safety Research 54, 49–54. doi:10.1016/j.jsr.2015.06.013

Gupta, A., Sharma, V., Ruparam, N.K., Jain, S., Alhammad, A., Ripon, M.A.K., 2014. Context-awareness based intelligent driver behavior detection: Integrating Wireless Sensor networks and Vehicle ad hoc networks. Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014 2155–2162. doi:10.1109/ICACCI.2014.6968230

Hammit, B.E., Ghasemzadeh, A., James, R.M., Ahmed, M.M., Young, R.K., 2018. Evaluation of weather-related freeway car-following behavior using the SHRP2 naturalistic driving study database. Transportation Research Part F: Traffic Psychology and Behaviour 59, 244–259. doi:10.1016/j.trf.2018.08.023

Harding, J., Powell, G., Yoon, R., Fikentscher, J., Doyle, C., Sade, D., Lukuc, M., Simons, J., Wang, J., 2014. Vehicle-to-Vehicle Communications : Readiness of V2V Technology for Application. Washington, DC. doi:Report No DOT HS 812 014

Harter, A., Hopper, A., Steggles, P., Ward, A., Webster, P., 2002. The anatomy of a context-aware application. Wireless Networks 8 2–3 , 187–197. doi:10.1023/A:1013767926256

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Second Edi. ed, Springer series in statistics. doi:10.1007/978-0-387-84858-7

Hayward, J.C., 1972. Near miss determination through use of a scale of danger. Highway Research Record 384, 24–34. doi:TTSC 7115

Hecht-Nielsen, R., 1989. Theory of the Back Propagation Neural Network. International Joint Conference on In Neural Networks 593–605. doi:10.1109/IJCNN.1989.118638

Henclewood, D., Rajiwade, S., 2015. Safety Pilot Model Deployment – Sample Data Environment Data Handbook 1.3.

Higgs, B., Abbas, M., 2015. Segmentation and clustering of car-following behavior: Recognition of driving patterns. IEEE Transactions on Intelligent Transportation Systems 16 1 , 81–90. doi:10.1109/TITS.2014.2326082

Hill, C., Elefteriadou, L., Kondyli, A., 2015. Exploratory Analysis of Lane Changing on Freeways Based on Driver Behavior. Journal of Transportation Engineering 141 4 , 04014090. doi:10.1061/(ASCE)TE.1943-5436.0000758

Hirst, S.J., Graham, R., 1997. The Format and Presentation of Collision Warnings. Ergonomics and safety of intelligent driver interfaces 203–219. doi:ISBN:0-8058-1956-8

Hoogendoorn, R.G., Breukink, H.J., Van Arem, B., 2012. A context aware intelligent speed adaptation system: A field operational test. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC 1091–1096. doi:10.1109/ITSC.2012.6338654

Hydén, C., 1987. The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique. Bulletin 70.

Jahangiri, A., Berardi, V.J., MacHiani, S.G., 2018. Application of Real Field Connected Vehicle Data for Aggressive Driving Identification on Horizontal Curves. IEEE Transactions on Intelligent Transportation Systems 19 7 , 2316–2324. doi:10.1109/TITS.2017.2768527

Jensen, M., Wagner, J., Alexander, K., 2011. Analysis of in-vehicle driver behaviour data for improved safety. International Journal of Vehicle Safety 5 3 , 197. doi:10.1504/IJVS.2011.042850

Johnson, D.A., Trivedi, M.M., 2011. Driving style recognition using a smartphone as a sensor platform. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC 1609–1615. doi:10.1109/ITSC.2011.6083078

Jolliffe, I.T., 2002. Principal Component Analysis, Springer Series in Statistics. Springer-Verlag, New York. doi:10.1007/b98835

Kenney, J.B., 2011. Dedicated short-range communications (DSRC) standards in the United States. Proceedings of the IEEE 99 7 , 1162–1182. doi:10.1109/JPROC.2011.2132790

Khodayari, A., Ghaffari, A., Kazemi, R., Braunstingl, R., 2012. A modified car-following model based on a neural network model of the human driver effects. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 42 6 , 1440–1449.

Kiefer, R., LeBlanc, D., Palmer, M., Salinger, J., Deering, R., Shulman, M., 1999. Development and Validation of Functional Definitions and Evaluation Procedures For Collision Warning/Avoidance Systems August , 75. doi:Report No. DOT-HS-808-964

Kim, T., Lovell, D., Park, Y., 2003. Limitations of Previous Models on Car-Following Behavior and Research Needs. Transportation Research Board, 82nd Annual Meeting 03 .

Kusano, K.D., Chen, R., Montgomery, J., Gabler, H.C., 2015. Population distributions of time to collision at brake application during car following from naturalistic driving data. Journal of Safety Research 54, 95–104. doi:10.1016/j.jsr.2015.06.011

Kusano, K.D., Montgomery, J., Gabler, H.C., 2014. Methodology for identifying car following events from naturalistic data. IEEE Intelligent Vehicles Symposium, Proceedings 281–285. doi:10.1109/IVS.2014.6856406

Lee, D., Yeo, H., 2016. Real-Time Rear-End Collision-Warning System Using a Multilayer Perceptron Neural Network. IEEE Transactions on Intelligent Transportation Systems 17 11

, 3087–3097. doi:10.1109/TITS.2016.2537878

Lee, S.E., Llaneras, E., Klauer, S.G., Sudweeks, J., 2007. Analyses of Rear-End Crashes and Near-Crashes in the 100-Car Naturalistic Driving Study to Support Rear-Signaling Countermeasure Development. Distribution October , 1–125.

Li, F., Hai Zhang, Huan Che, Xiaochen Qiu, 2016. Dangerous driving behavior detection using smartphone sensors, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC). IEEE, pp. 1902–1907. doi:10.1109/ITSC.2016.7795864

Lin, Q., Feng, R., Cheng, B., Lai, J., Zhang, H., Mei, B., 2008. Analysis of Causes of Rear-end Conflicts Using Naturalistic Driving Data Collected by Video Drive Recorders. SAE Technical Paper 2008 724 . doi:10.4271/2008-01-0522

Loke, S.W., 2006. Context-Aware Pervasive Systems: Architectures for a New Breed of Applications 0849372550 , 240.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics 5 4 , 115–133. doi:10.1007/BF02478259

McFadden, J., Yang, W.T., Durrans, S., 2001. Application of artificial neural networks to predict speeds on two-lane rural highways. Transportation Research Record: Journal of the Transportation Research Board 1751, 9–17.

Mclaughlin, S., Hankey, J., Dingus, T., 2009. Driver Measurement : Methods and Applications, in: International Conference on Engineering Psychology and Cognitive Ergonomics. pp. 404–413.

McLaughlin, S., Hankey, J.M., Dingus, T. a, Klauer, S.G., 2009. Development of an FCW algorithm evaluation methodology with evaluation of three alert algorithms. National Highway Traffic Safety Administration, DOT HS 811 June , 145.

McLaughlin, S., Hankey, J.M., Dingus, T.A., 2008. A method for evaluating collision avoidance systems using naturalistic driving data. Accident Analysis and Prevention 40 1 , 8–16. doi:10.1016/j.aap.2007.03.016

Meseguer, J.E., Calafate, C.T., Cano, J.C., Manzoni, P., 2013. DrivingStyles: A smartphone application to assess driver behavior. Proceedings - International Symposium on Computers and Communications 535–540. doi:10.1109/ISCC.2013.6755001

Mohammed, M., Ke, Y., Gao, J., Zhang, H., El-Basyouny, K., Qiu, T.Z., 2016. Connected vehicle V2I communication application to enhance driver awareness at signalized intersections. Proceedings, Annual Conference - Canadian Society for Civil Engineering 4.

Montgomery, J., Kusano, K.D., Gabler, H.C., 2014. Age and gender differences in time to collision at braking from the 100-car naturalistic driving study. Traffic injury prevention 15 Suppl 1 November , S15-20. doi:10.1080/15389588.2014.928703

Naujoks, F., Kiesel, A., Neukum, A., 2016a. Cooperative warning systems: The impact of false and unnecessary alarms on drivers??? compliance. Accident Analysis and Prevention 97, 162–175. doi:10.1016/j.aap.2016.09.009

Naujoks, F., Purucker, C., Neukum, A., 2016b. Secondary task engagement and vehicle automation - Comparing the effects of different automation levels in an on-road experiment. Transportation Research Part F: Traffic Psychology and Behaviour 38, 67–82. doi:10.1016/j.trf.2016.01.011

Neuman, T., Pfefer, R., Slack, K., Hardy, K., Raub, R., Lucke, R., Wark, R., 2003. Guidance for implementation of the AASHTO strategic highway safety plan - Volume I: A guide for addressing aggressive-driving collisions. Wcha Net 1, 75.

Osafune, T., Takahashi, T., Kiyama, N., Sobue, T., Yamaguchi, H., Higashino, T., 2017. Analysis of Accident Risks from Driving Behaviors. International Journal of Intelligent Transportation Systems Research 15 3 , 192–202. doi:10.1007/s13177-016-0132-0

Peng, J., Guo, Y., Fu, R., Yuan, W., Wang, C., 2015. Multi-parameter prediction of drivers' lane-changing behaviour with neural network model. Applied Ergonomics 50, 207–217. doi:10.1016/j.apergo.2015.03.017

Perez, M.A., Sudweeks, J.D., Sears, E., Antin, J., Lee, S., Hankey, J.M., Dingus, T.A., 2017. Performance of basic kinematic thresholds in the identification of crash and near-crash events within naturalistic driving data. Accident Analysis and Prevention 103, 10–19. doi:10.1016/j.aap.2017.03.005

Purucker, C., Schneider, N., Rüger, F., Frey, A., 2018. Validity of Research Environments – Comparing Criticality Perceptions Across Research Environments, in: UR:BAN Human Factors in Traffic. Springer Fachmedien Wiesbaden, Wiesbaden, pp. 423–446. doi:10.1007/978-3-658-15418-9_25

R.Stollof, E., 2008. Intersection and Junction Fatalities in the Context of Access Management. 8th National Conference on Access Management July .

Rakha, H., 2009. Validation of Van Aerde's simplified steadystate car-following and traffic stream model. Transportation Letters 1 3 , 227–244. doi:10.3328/TL.2009.01.03.227-244

Rakha, H., Crowther, B., 2003. Comparison and calibration of FRESIM and INTEGRATION steady-state car-following behavior. Transportation Research Part A: Policy and Practice 37 1 , 1–27. doi:10.1016/S0965-8564(02)00003-4

Rodriguez Gonzalez, A.B., Wilby, M.R., Vinagre Diaz, J.J., Sanchez Avila, C., 2014. Modeling and detecting aggressiveness from driving signals. IEEE Transactions on Intelligent Transportation Systems 15 4 , 1419–1428. doi:10.1109/TITS.2013.2297057

Rong, J., Mao, K., Ma, J., 2011. Effects of Individual Differences on Driving Behavior and Traffic Flow Characteristics. Transportation Research Record: Journal of the Transportation Research Board 2248, 1–9. doi:10.3141/2248-01

Ruscio, D., Ciceri, M.R., Biassoni, F., 2015. How does a collision warning system shape driver's brake response time? the influence of expectancy and automation complacency on real-life emergency braking. Accident Analysis and Prevention 77, 72–81. doi:10.1016/j.aap.2015.01.018

SAE, 2013. Dedicated Short Range Communications (DSRC) Message Set Dictionary 4970, 724–776.

SAE International, 2014. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE International J3016, 1–12. doi:10.4271/J3016_201609

Saifuzzaman, M., Zheng, Z., 2014. Incorporating human-factors in car-following models : A review of recent developments and research needs. Transportation Research Part C 48, 379–403. doi:10.1016/j.trc.2014.09.008

Sarvi, M., 2013. Heavy commercial vehicles-following behavior and interactions with different vehicle classes. Journal of Advanced Transportation 47 6 , 572–580. doi:10.1002/atr.182

Sathyanarayana, A., Boyraz, P., Hansen, J.H.L., 2011. Information fusion for robust "context and driver aware" active vehicle safety systems. Information Fusion 12 4 , 293–303. doi:10.1016/j.inffus.2010.06.004

Satzoda, R.K., Trivedi, M.M., 2015. On Enhancing Lane Estimation using Contextual Cues 8215 c , 1–12. doi:10.1109/TCSVT.2015.2406171

Sayed, T., Abdelwahab, W., Navin, F., 1995. Identifying Accident-Prone Locations Using Fuzzy

Pattern Recognition. Journal of Transportation Engineering 121 4 , 352–358. doi:10.1061/(ASCE)0733-947X(1995)121:4(352)

Schilit, B., Adams, N., Want, R., 1994. Context-aware computing applications. IEEE W rkshop on Mobile Computing Systems and Applications 1–7. doi:10.1109/MCSA.1994.512740

Singh, G., Bansal, D., Sofat, S., 2017. A smartphone based technique to monitor driving behavior using DTW and crowdsensing. Pervasive and Mobile Computing 40, 56–70. doi:10.1016/j.pmcj.2017.06.003

Smiley, A., 1996. Overview and methodologies of the ITS safety evaluation process, in: ITS America and NHTSA, Proc. Workshop on Safety Evaluation of Intelligent Transport Systems. ITS America, pp. 5–20.

Staubach, M., Schebitz, N., Köster, F., Kuck, D., 2014. Evaluation of an eco-driving support system. Transportation Research Part F: Traffic Psychology and Behaviour 27 PA , 11–21. doi:10.1016/j.trf.2014.09.006

Sundbom, M., Falcone, P., Sjoberg, J., 2013. Online driver behavior classification using probabilistic ARX models. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC 1107–1112. doi:10.1109/ITSC.2013.6728380

Tang, K., Chen, S., Liu, Z., Khattak, A.J., 2018. A tensor-based Bayesian probabilistic model for citywide personalized travel time estimation. Transportation Research Part C: Emerging Technologies 90, 260–280. doi:10.1016/j.trc.2018.03.004

Tang, T., Wang, Y., Yang, X., Wu, Y., 2012. A new car-following model accounting for varying road condition. Nonlinear Dynamics 70 2 , 1397–1405. doi:10.1007/s11071-012-0542-8

Tawfeek, M.H., El-Basyouny, K., 2018a. Network-level Comparison of Various Forward Collision Warning Algorithms. Simulation: Transactions of the Society for Modeling and

Simulation International.

Tawfeek, M.H., El-Basyouny, K., 2018b. A perceptual forward collision warning model using naturalistic driving data. Canadian Journal of Civil Engineering 45 10 , 899–907. doi:10.1139/cjce-2017-0592

Tawfeek, M.H., El-Basyouny, K., 2018c. Using Naturalistic Driving Data to Quantify Driver Following Behavior During Braking. Proceedings, Annual Conference - Canadian Society for Civil Engineering.

Tawfeek, M.H., El Esawey, M., El-Araby, K., Abdel-Latif, H., 2018. Calibration and Validation of Micro-Simulation Models Using Measurable Variables, in: Proceedings, Annual Conference - Canadian Society for Civil Engineering. Fredericton, NB, Canada.

The Mathworks Inc., 2016. MATLAB - MathWorks. Www.Mathworks.Com/Products/Matlab. doi:2016-11-26

Tijerina, L., Garrott, W., 1997. A Reliability Theory Approach to Estimate the Potential Effectiveness of a Crash Avoidance System to Support Lane Change Decisions. SAE Technical Paper 970454.

Transport Canada, 2018. National Collision Database [WWW Document]. Transport Canada. URL http://open.canada.ca/data/en/dataset/1eb9eba7-71d1-4b30-9fb1-30cbdab7e63a (accessed 8.10.18).

Treat, J., Tumbas, N., McDonald, S., 1979. Tri-Level Study of the Causes of Traffic Accidents. Executive Summary. Vision Research 42 21 , 2419–2430. doi:00771862

Treiber, M., Hennecke, A., Helbing, D., 2000. Congested traffic states in empirical observations and microscopic simulations. Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics 62 2 , 1805–1824. doi:10.1103/PhysRevE.62.1805

UDRIVE, n.d. European Naturalistic Driving Study The First Naturalistic Driving Study [WWW Document]. 2012. URL http://www.udrive.eu/index.php/about-udrive/what-is-naturalistic-driving (accessed 1.1.17).

USDOT, 2018. ITS Public Data Hub [WWW Document]. URL https://www.its.dot.gov/data/ (accessed 1.3.18).

Vahdat-Nejad, H., Ramazani, A., Mohammadi, T., Mansoor, W., 2016. A survey on context-aware vehicular network applications. Vehicular Communications 3, 43–57. doi:10.1016/j.vehcom.2016.01.002

van Driel, C.J.G., Hoedemaeker, M., van Arem, B., 2007. Impacts of a Congestion Assistant on driving behaviour and acceptance using a driving simulator. Transportation Research Part F: Traffic Psychology and Behaviour 10 2 , 139–152. doi:10.1016/j.trf.2006.08.003

Vogel, K., 2003. A comparison of headway and time to collision as safety indicators. Accident Analysis and Prevention 35 3 , 427–433. doi:10.1016/S0001-4575(02)00022-2

Wada, T., Doi, S., Imai, K., Tsuru, N., Isaji, K., Kaneko, H., 2007. Analysis of Braking Behaviors in Car Following Based on A Performance Index for Approach and Alienation. 2007 IEEE Intelligent Vehicles Symposium 724 , 776–790. doi:10.1109/IVS.2007.4290172

Wan, J., Zhang, D., Zhao, S., Yang, L., Lloret, J., 2014. Context-aware vehicular cyber-physical systems with cloud support: Architecture, challenges, and solutions. IEEE Communications Magazine 52 8 , 106–113. doi:10.1109/MCOM.2014.6871677

Wang, J., Lu, M., Li, K., 2010a. Characterization of Longitudinal Driving Behavior by Measurable Parameters. Transportation Research Record: Journal of the Transportation Research Board 2185, 15–23. doi:10.3141/2185-03

Wang, J., Lu, X., Xiao, Q., Lu, M., 2010b. Comparison of Driver Classification Based on

Subjective Evaluation and Objective Experiment. Transportation Research Record 224 July , 1–13.

Wang, J., Zhang, L., Zhang, D., Li, K., 2013. An adaptive longitudinal driving assistance system based on driver characteristics. IEEE Transactions on Intelligent Transportation Systems 14 1 , 1–12. doi:10.1109/TITS.2012.2205143

Wang, W., Xi, J., Chong, A., Li, L., 2017. Driving Style Classification Using a Semisupervised Support Vector Machine. IEEE Transactions on Human-Machine Systems 47 5 , 650–660. doi:10.1109/THMS.2017.2736948

Wang, W., Xi, J., Zhao, D., 2018. Learning and Inferring a Driver's Braking Action in Car-Following Scenarios. IEEE Transactions on Vehicular Technology 67 5 , 3887–3899. doi:10.1109/TVT.2018.2793889

Wang, X., Abdel-Aty, M., Nevarez, A., Santos, J.B., 2009. Investigation of Safety Influence Area for Four-Legged Signalized Intersections: Nationwide Survey and Empirical Inquiry. Transportation Research Record: Journal of the Transportation Research Board 2083 1 , 86–95. doi:10.3141/2083-10

Wang, X., Zhu, M., Chen, M., Tremont, P., 2016. Drivers' rear end collision avoidance behaviors under different levels of situational urgency. Transportation Research Part C: Emerging Technologies 71, 419–433. doi:10.1016/j.trc.2016.08.014

Wang, Y., Jiang, J., Mu, T., 2013. Context-aware and energy-driven route optimization for fully electric vehicles via crowdsourcing. IEEE Transactions on Intelligent Transportation Systems 14 3 , 1331–1345. doi:10.1109/TITS.2013.2261064

Wege, C., Will, S., Victor, T., 2013. Eye movement and brake reactions to real world brake-capacity forward collision warnings - A naturalistic driving study. Accident Analysis and

Prevention 58, 259–270. doi:10.1016/j.aap.2012.09.013

WHO, 2013. Global status report on road safety 2013, Who.

Williams, V.H., McLaughlin, S., Williams, S.L., Buche, T., 2015. Exploratory Analysis of Motorcycle Incidents Using Naturalistic Riding Data. Transportation Research Record: Journal of the Transportation Research Board 2520, 151–156. doi:10.3141/2520-17

Wu, C., Sun, C., Chu, D., Huang, Z., Ma, J., Li, H., 2016. Clustering of Several Typical Behavioral Characteristics of Commercial Vehicle Drivers Based on GPS Data Mining. Transportation Research Record: Journal of the Transportation Research Board 2581, 154–163. doi:10.3141/2581-18

Wu, K.-F., Thor, C.P., 2015. Method for the Use of Naturalistic Driving Study Data to Analyze Rear-End Crash Sequences. Transportation Research Record: Journal of the Transportation Research Board 2518, 27–36. doi:10.3141/2518-04

Wu, Y., Nakamura, H., Asano, M., 2013. A comparative study on crash-influencing factors by facility types on urban expressway. Journal of Modern Transportation 21 4 , 224–235. doi:10.1007/s40534-013-0024-9

Yang, B., Guo, C., Ma, Y., Jensen, C.S., 2015. Toward personalized, context-aware routing. VLDB Journal 24 2 , 297–318. doi:10.1007/s00778-015-0378-1

Yang, H., Kitamura, R., Jovanis, P.R., Vaughn, K.M., Abdel-Aty, M.A., 1993. Exploration of route choice behavior with advanced traveler information using neural network concepts. Transportation 20, 199–223. doi:10.1007/BF01307059

Yang, L., Ma, R., Zhang, H.M., Guan, W., Jiang, S., 2018. Driving behavior recognition using EEG data from a simulated car-following experiment. Accident Analysis & Prevention 116, 30–40. doi:10.1016/J.AAP.2017.11.010

Yang Zheng, Wang, J., Xiaofei Li, Chenfei Yu, Kodaka, K., Li, K., 2014. Driving risk assessment using cluster analysis based on naturalistic driving data. 17th International IEEE Conference on Intelligent Transportation Systems (ITSC) OCTOBER , 2584–2589. doi:10.1109/ITSC.2014.6958104

Ye, M., Osman, O.A., Ishak, S., Hashemi, B., 2017. Detection of driver engagement in secondary tasks from observed naturalistic driving behavior. Accident Analysis and Prevention 106, 385–391. doi:10.1016/j.aap.2017.07.010

Young, K.L., Regan, M.A., Triggs, T.J., Tomasevic, N., Stephan, K., Mitsopoulos, E., 2007. Impact on car driving performance of a following distance warning system: Findings from the Australian transport accident commission SafeCar project. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 11 3 , 121–131. doi:10.1080/15472450701410411

Zhao, Y., Wagh, A., Hou, Y., Hulme, K., Qiao, C., Sadek, A.W., 2016. Integrated Traffic-Driving-Networking Simulator for the Design of Connected Vehicle Applications: Eco-Signal Case Study. Journal of Intelligent Transportation Systems 20 1 , 75–87. doi:10.1080/15472450.2014.889920

Zheng, Y., Chase, R.T., Elefteriadou, L., Sisiopiku, V., Schroeder, B., 2017. Driver types and their behaviors within a high level of pedestrian activity environment. Transportation Letters 9 1 , 1–11. doi:10.1080/19427867.2015.1131943