Technical University of Munich

Department of Civil, Geo and Environmental Engineering

Chair of Transportation Systems Engineering

MASTER'S THESIS

Identification of Dangerous Driving Behaviour Using Naturalistic Driving Data

Ariful Hoque Shaon

Supervised by

Dr. Christos Katrakazas

March 28, 2019

Abstract

Road accidents are one of the most predominant factors for deaths throughout the world. With the inclusion of several driver assistance systems, intelligent vehicles are becoming peak of the automotive industry to mitigate accidents. Autonomous vehicles are widely considered safer, because of the introduction of advanced robotics and Advanced Driver Assistance Systems (ADAS) into the task of driving. However, the main challenge for AVs is to properly detect dangerous situations and react properly to avoid potential collisions. To overcome this challenge, it is important to assess current traffic situation and vehicle dynamics for real-time collision prediction. This thesis provides an insight to identify and predict dangerous driving behaviour for autonomous vehicles in an uncontrolled intersection for rear-end collision scenarios. A large naturalistic driving dataset containing single vehicle data of position, speed and heading is analyzed to predict future conflicts by utilizing machine learning classification techniques.

To that aim, vehicle level data are collected using sensors installed on a vehicle, which deliberately passes through an uncontrolled T-intersection. The vehicle passed approximately ten times in each of the six possible manoeuvres. A circular area of interest with radius of 35 meters is selected around the center of intersection. Based on this bounding area, vehicle trajectories are extracted from position data based on their entry and exit points. Trajectories are then time-shifted, so as to imitate interactions among them and develop rear-end collision scenarios. Finally, Time-to-Collision (TTC) is used as a surrogate safety indicator to identify dangerous behaviour.

A total of 11,208 gap observations are counted in all six manoeuvres in the bounding area. Among them 35.96% observations are marked as dangerous, where TTC lie below the threshold value of 1.5 seconds. It is observed that TTC gets lower when the vehicle approaches to the intersection. Moreover, there is an inverse relationship between TTC and speed difference. High difference of speed between the following vehicle and lead vehicle leads to lower TTC and results in dangerous situation. On the contrary, low speed difference shows high TTC and low collision risk. It is observed that TTC decreases exponentially with increase in speed difference between the following and lead vehicle.

Finally, different machine learning classifiers are tested to classify and predict dangerous situations considering speed difference as the independent variable or predictor. After analyzing performance matrices, it is observed that Random Forest (RF) performs better than other classifiers in terms of different performance matrices and gives a lower rate of false alarm (less than 7%). Area under the ROC curve also increases for RF. Later on, RF classifier is employed in all the six manoeuvres to classify dangerous driving behaviour. However, in some manoeuvres, it gives higher false prediction due to the high imbalance between safe and collision-prone test samples. It is expected that more sophisticated real-world traffic data and integration of more advanced classification techniques like imbalanced learning or deep learning are more likely to give better prediction of collisions.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisor Dr. Christos Katrakazas of the Chair of Transportation Systems Engineering at Technical University of Munich. His constant support and guidance with enormous enthusiasm throughout the thesis has an immense effect on the completion of this research.

I would like to thank and dedicate this thesis to my beloved wife Sharna, my parents and all the friends for their unfailing support and endless love.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADAS   Advanced Driver Assistance Systems

AUC   Area Under the ROC Curve

AV   Autonomous Vehicle

CV   Cross-Validation

DBN   Dynamic Bayesian Network

ITS   Intelligent Transportation Systems

kNN   k-Nearest Neighbors

MDP   Markov Decision Process

MOMDP   Mixed observability Markov Decision Process

NDS   Naturalistic Driving Study

NLCP   Network-level collision prediction

PET   Post-Encroachment Time

POMDP   Partially Observable Markov Decision Process

RF   Random Forest

ROC   Receiver Operating Characteristic

RVM   Relevance Vector Machine

SSAM   Surrogate Safety Assessment Model

SVM   Support Vector Machine

TA   Time-to-Accident

THW   Time-to-Headway

TTC   Time-to-Collision

V2I   Vehicle-to-Infrastructure

V2V   Vehicle-to-Vehicle

WHO   World Health Organization

1 Introduction

1.1 Background and motivation

Road accidents have become one of the most increasing factors for deaths in the present world. According to WHO (2015), more than 1.25 million people died around the world in 2013 because of road accidents and they are the biggest non-health reason for deaths. WHO also predicted that, road traffic injuries will become the seventh leading cause of human death by the year 2030.

As the severity of losses is immense due to road accident, traffic safety issue has got a lot of attention in the research industry over the years. Evans (2004) defined 'traffic safety' as "*the absence of unintended harm to living creatures or inanimate objects*". Treat et al. (1979) stated that road environment, vehicle state and human factors all contribute to collisions. According to Oh et al. (2001) traffic dynamics also play a significant role in conflict occurrence along with these three components. The main challenge in road safety is to detect or predict dangerous situation and react properly to avoid collision. This is the reason why, identification of dangerous driving behaviour and real-time collision prediction have gained major focus in the field of Intelligent Transportation Systems (ITS).

Abdel-Aty and Pande (2005) stated that real-time collision prediction can be performed by estimating the probability of a collision occurrence from traffic data for a short-time prediction horizon. ITS has facilitated the research of collision prediction immensely by increasing the availability of traffic data. Real-time collision prediction has become easier with the inclusion of advanced loop detectors and sensors. Advancement in smartphone technology has also helped immensely in cost effective and additional traffic data collection (Guido et al. 2012, Vlahogianni and Barmpounakis 2017). Image and video processors have been used in many research successfully for collecting less noisy data for collision detection (Ikeda et al. 1999, Astarita et al. 2011). Automatic Vehicle Identification (AVI) technology is also practiced in literature in recent times for collision data collection (Yu et al. 2013, Ahmed et al. 2012).

Apart from growing research in road safety, accidents are still predominant. According to the Bureau of Transportation Statistics (2015), in USA around 32,000 people are died, and more than two millions are injured in road accidents every year. According to the statistics from NHTSA (National Highway Traffic Safety Administration and U.S. Department of Transportation) (2011), about 41% road accidents are done by drunk driving, 10% by distraction of drivers and 2.5% by fatigue. Several studies also showed that more than 90% traffic accidents were caused by driving mistakes or human errors (Staubach 2009, Singh 2015, Paden et al. 2016). Human error can be caused due to lack of information or misuse of information like misinterpretation of other vehicle's relative speed or distance. Driver's age and mental stage also play vital role in the occurrence of traffic collision.

ITS research has been rapidly increasing to minimize human error in driving. With the inclusion of several driver assistance systems, "Intelligent Vehicles" have become a big step in the automotive industry to mitigate road accidents. Autonomous vehicles (AVs) increase driving safety significantly by eliminating human errors like drunk driving, fatigue, distraction etc. Introduction of advanced robotics and Advanced Driver Assistance Systems (ADAS) in AVs have added better perception and decision-making when driving on roads. Many authors have stated the benefits of autonomous vehicles for increasing road safety. Fagnant and Kockelman (2015) described the advantages of autonomous vehicles in terms of increasing safety, solving for congestion, decreasing travel time and cheaper travelling. According to them, parking cost and space will be reduced, and car sharing can be introduced for less traffic. Autonomous vehicles will play a great role in the near future for solving mobility problems by reaching to different groups of people (Sivak and Schoettle 2015). Low carbon mobility is another environmental advantage of AVs to reduce

emissions (Thomopoulos and Givoni 2015). Introduction of vehicle-to-vehicle (V2V) communication and 3D laser sensor technology will inform drivers about the collision happened on road and warn them, which may prevent after-collision congestion (Wei et al. 2013, Jiménez et al. 2016). In the study of the A.T. Kearney (2016), it is expected that road accidents will be reduced by 70% by introducing autonomous driving. Also, the annual savings in USA will be 1.3 trillion US dollars due to the introduction of AVs.
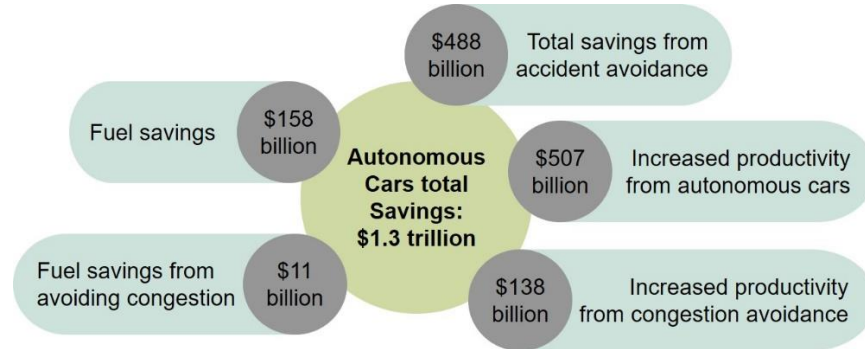
Figure 1.1 Expected annual economic benefits of autonomous vehicles in USA (A.T. Kearney 2016)

In recent years, enormous number of researches are going on to enhance the safety of autonomous vehicles after realizing the benefits associated with AVs in future. The challenge of adapting with surrounding environment (human driven vehicles, pedestrians etc.) while driving on complex road geometry is a big focus in safety assessment of autonomous driving. Moreover, future collision prediction from present traffic data has got a lot of attention in road safety research along with the research of identifying collision-prone road geometry.

1.2   Problem statement

As autonomous vehicle is the most recent advancement in automotive industry, the underlying safety associated with it forms a big problem. Safe navigation of AVs with other traffic participants, as well as threats and challenges faced by different impact groups due to the introduction of AVs are studied intensively. However, perfect safety assessment of AVs is not achieved yet.

"Perfect safety is really an impossible goal. It's really about improving the probability of safety – that's the only thing possible."- (Lambert 2016)

Manoeuvre planning or path planning is the fundamental of risk assessment for autonomous vehicles (Katrakazas et al. 2015). Typically, a motion model predicts trajectories of other vehicles and estimates the collision risk associated with the trajectory of autonomous vehicle. However, computational complexity is emerged in different complex road geometries for getting a safe trajectory, where traffic participants are assumed to navigate independently (Lefèvre et al. 2014). Research is going on to integrate human interaction or driving behaviour without predicting other vehicle's trajectory and information of traffic data in the risk assessment of autonomous vehicles (Agamennoni et al. 2012, Gindele et al. 2015, Lefevre et al. 2012). But in reality, inter-vehicle communication is not always feasible. Therefore, it is often assumed when modelling collision risk prediction (Paden et al. 2016). There is still a need for robust collision prediction model for the risk assessment of autonomous vehicles.

Real-time collision prediction has become popular in risk assessment in the recent years. According to Katrakazas (2017), Real-time collision prediction for autonomous vehicles is typically hierarchically structured into the following four steps.

a) Identifying traffic variable/variables as predictors

b) Traffic data collection for both normal and collision-prone situations

c) Predict collision by classification technique

d) Performance evaluation of the classification model

But most of the times noisy or missing traffic data lead to a bad prediction by the classification model (Xu et al. 2015). Moreover, an imbalance between data emerges due to the lack of collision data compared to safe situations. Xu et al. (2016) stated that this imbalance leads to biased classification and high number of false alarms. Most of the literatures have neglected this effect of imbalance. According to Roshandel et al. (2015), the ratio of collision-prone and safe cases should not exceed 1:5 for better prediction performance of the classifiers. Inclusion of more disaggregated data is also important while building a classification model (Katrakazas 2017).

This thesis will attempt to identify dangerous driving behaviour based on naturalistic driving data with respect to the performed manoeuvre. A large naturalistic dataset will be used for this approach and machine learning classification techniques will be employed for collision risk assessment.

1.3   Aim and objectives

This thesis aims to build a vehicle-level collision risk prediction model that can identify dangerous driving behaviour using naturalistic driving data. The following objectives have been formulated, so as to fulfil this aim:

- To review existing motion planning and real time collision prediction models from literature.
- To extract car-following events from time shifted trajectories to identify rear-end collisions.
- To implement machine learning classification techniques for predicting dangerous driving behaviour leading to collision.
- To evaluate classifier performance and build a prediction model to predict dangerous driving behaviour associated with autonomous vehicles.

1.4   Expected contributions

This thesis contributes to the knowledge of methodological and practical level in the following ways:

- Existing approaches of real-time collision prediction are mostly based on network level traffic data. Risk assessment can be enhanced by predicting real-time collision based on vehicle dynamics or driving behavior named vehicle level collision prediction. This thesis uses a single vehicle data from a naturalistic dataset to predict future collision situations due to dangerous driving, which supports the argument that collision prediction model does not need to depend fully on trajectory studies of surrounding vehicles in a road segment.
- Concept of time shifted trajectories is implemented to extract car-following events from a single vehicle data of the naturalistic driving study. Generating trajectories for actual collision is not feasible. Hence, controlling the time shifting of trajectories can be a good alternative for collision prediction and safety assessment.
- Dangerous driving behaviour is identified for rear-end collision around an uncontrolled T-intersection by setting a threshold value of surrogate measure Time-to-Collision (TTC). For classifying dangerous behaviour, speed difference is considered as the predictor instead of the individual speed of vehicles. With the advancement in vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, avoiding dangerous situations will be easier by the early prediction of collision-prone situations due to dangerous driving behaviour obtained from the classification technique used in this thesis.

1.5   Overview of the thesis

This report contains six main chapters. Chapter 1 introduces the background and motivation of the research and problem definition. It also describes the main objectives of the thesis and expected contributions to the field of safety assessment of AVs. In Chapter 2, brief literature of AV motion planning approaches and safety assessment is reviewed. The same chapter also provides a brief description of different surrogate safety indicators used in collision risk prediction models with related work. Chapter 3 describes the data collection procedure, giving insight on the data used for the research and a visualization of the study area. Chapter 4 presents the methodology followed throughout the thesis in order to solve the research problem. Chapter 5 contains the analysis results obtained from the methods applied and represents a driving behaviour prediction model using machine learning classification techniques with corresponding performance evaluation. The final chapter provides a summary of the report as well as limitations and possible future work of the thesis. An overview of the thesis is illustrated in Figure 1.2.
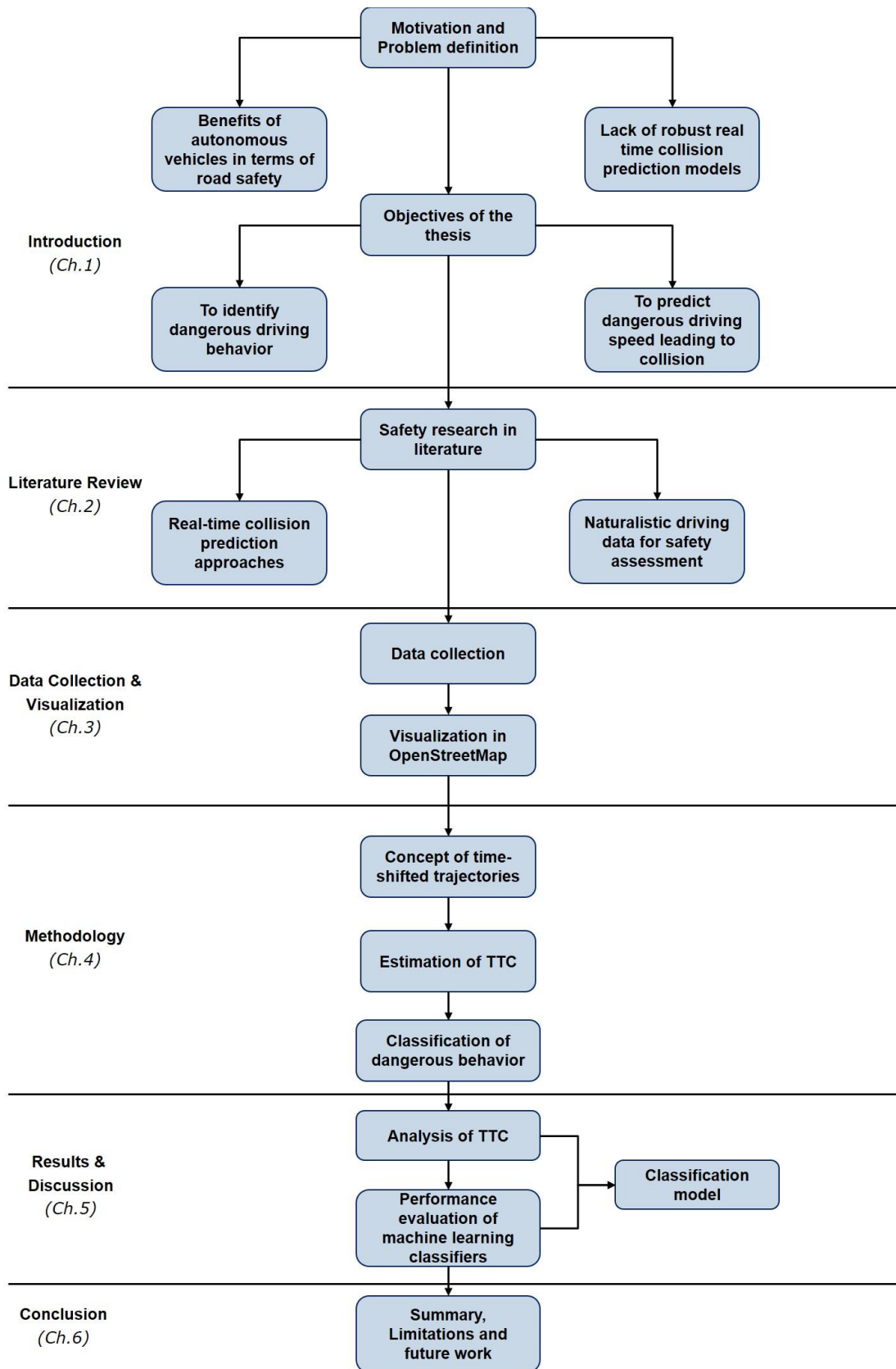
Figure 1.2 Thesis overview

2 Literature Review

2.1 Planning for autonomous vehicles

Research on autonomous vehicles has become the first choice in automotive industry after the success of DARPA (Defense Advanced Research Projects Agency) Grand challenge (Buehler et al. 2009, Stilgoe 2017, Schwarting et al. 2018, Kala and Warwick 2013). Different automotive companies like Google (Dolgov 2016), Tesla (Stilgoe 2017), BMW (Ziegler et al. 2014) have initiated their research to make autonomous vehicles. It is expected in current literature that fully autonomous vehicles will appear within the next decade and a huge amount of AVs will be driven on road within next 50 years (Hörl et al. 2016). Litman (2015) also predicted AVs to be affordable for the majority of people between 2040 and 2060. However, to replace conventional cars and making autonomous vehicles accepted to all groups of people, it is needed to establish public trust in safety associated with AVs. Research is going on for increasing safety level of AVs, although number of accidents have occurred (Dolgov 2016, Stilgoe 2017). Therefore, expected safety of AVs is still not achieved (Fagnant and Kockelman 2015).

Reliable and robust planning is a prerequisite for safe navigation of autonomous vehicles. The main challenge of autonomous vehicles while driving on the road is decision making on critical situations. Planning algorithms for decision making are key to navigate the vehicle towards a safe manoeuvre and reach the destination bound by traffic rules and road boundaries (Zhang et al. 2013). Varaiya (1993) suggested that planning for on-road autonomous vehicles is hierarchically structured into four parts named a) route planning, b) path planning, c) manoeuvre choice and d) trajectory planning. Paden et al. (2016) integrated these four steps in decision making process and visualized hierarchically as shown in Figure 2.1. Route planning is defined as finding the best route to reach the destination and it is not associated with vehicle dynamics or traffic interaction. Behavioural layer (path planning), Motion planning (manoeuvre choice) and local feedback control or vehicle control (trajectory planning) are associated with the vehicle dynamics, road geometry and traffic interactions.

The road environment and state space of the vehicle (position, speed, orientation) need to be represented for searching the path and planning. It can be done by visualizing the sensor data in a digital map. The digital representation of the road environment must be performed with efficiency, density and expressiveness (Howard 2009). Widely used representing techniques in literature are Voronoi diagrams (Dolgov et al. 2010, Lee et al. 2014), driving corridors (Wille et al. 2010, Jeon et al. 2013, Hardy and Campbell 2013), occupancy grids (Kammel et al. 2009, Hundelshausen et al. 2008, Zhao et al. 2011, Xu et al. 2014), cost maps (Schroder et al. 2008, Broggi et al. 2012, Murphy and Newman 2011) and state lattices (Pivtoraiko and Kelly 2005, McNaughton et al. 2011, Ziegler and Stiller 2009). These representation techniques are sometimes combined with other techniques for better planning of path, manoeuvre and trajectory. For instance, potential fields were combined with Voronoi diagrams to generate Voronoi fields to obtain better results by Dolgov et al. (2010).
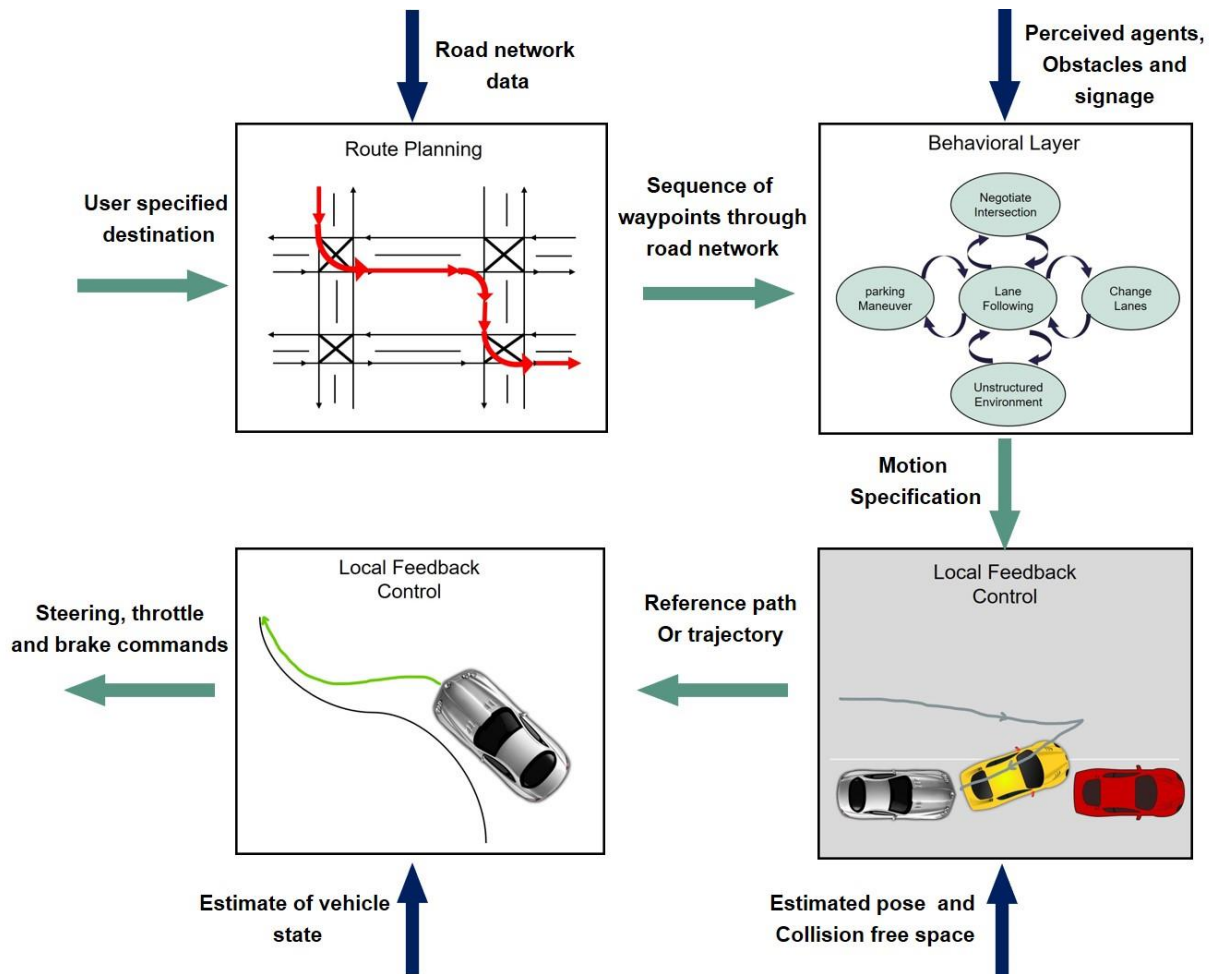
Figure 2.1 Hierarchy of decision-making process (Paden et al. 2016)

After the search space representation, planning algorithms are employed for finding the best path, manoeuvre and trajectory. According to Katrakazas et al. (2015), planning of on-road driving of autonomous vehicles is divided into three parts:

1) Search for the best path to follow

2) Search for the best manoeuvre to perform and

3) Search for the best trajectory

Path-planning or behavioural decision making comes after finding a route. Path is the continuous sequence of configurations from start to end with the boundary configurations (Eskandarian 2012). Path-planning refers to finding the best geometric path in the selected route, without any collision with traffic participants or obstacles while driving according to road boundary and traffic rules.

After finding the best path, best manoeuvre should be performed. Manoeuvre indicates the motion of the vehicle regarding vehicle state (e.g. position and speed). Manoeuvre planning refers to the best driving behaviour to be performed in the current context like going straight, changing lane, overtaking, turning etc. (Paden et al. 2016). Autonomous vehicles need to interact with surrounding traffic participants before performing the best and safest manoeuvre. According to Katrakazas (2017), Manoeuvre planning can be

performed by two categories named motion planning (also obstacle prediction) and decision-making module.

Trajectory represents the states or points followed by vehicle in a manoeuvre. Trajectory planning is referred by finding the transition of a vehicle from one state to another in real-time, constrained by road boundaries and traffic rules. Trajectory planning is often addressed as motion planning and is parameterized by time and velocity or acceleration (Katrakazas et al. 2015). Trajectory planning can be done at regular intervals of time depending on the position data, which can be obtained from sensor measurements. Risk assessment can be performed by studying future trajectories that can collide or detecting unexpected behaviour or manoeuvres by traffic participants (Lefèvre et al. 2014, Ammoun and Nashashibi 2009).

2.1.1 Motion planning models for manoeuvre choice

A key challenge in predicting collision is dealing with uncertainties like inaccurate or unclear motion intentions of vehicles (Song et al. 2016). Gadepally (2013) used hierarchical hidden Markov models and finite state machines to predict the motion of vehicles at intersections. However, his approach needs large amount of data for training. Armand et al. (2014) used ontology to predict behaviour of the vehicles for limited manoeuvres. Alin et al. (2012) predicted vehicle trajectories as spline functions by using grid-based Bayesian filter, although they only considered cut-in and lane changing manoeuvres.

Vehicle level collision risk prediction of autonomous vehicles are based on different motion models which describe vehicle's movement considering the surrounding (Katrakazas et al. 2015). Lefèvre et al. (2014) surveyed different motion models and prediction approaches for the risk assessment of intelligent vehicles. They classified motion models into three categories named physics-based, maneuver-based and interaction-aware motion models. Physics -based motion models are widely used in literature to predict future motion or trajectory using dynamic and kinematic models. These models follow the laws of physics and face the limitation of predicting motion of a very short time. They are also incapable of anticipating any change in motion for abrupt braking or acceleration. Maneuver-based motion models are based on predicting the intention of vehicles by clustering the trajectories or estimating manoeuvres. These models also do not consider traffic environment into account and do not adapt well in every road layout like road intersections because each motion model is trained for specific road geometry. The last one is interaction-aware motion models which are based on inter-vehicle dependencies and are very handy in collision risk prediction. These models perform well in predicting collision risk at different road geometry like intersections, where priority rules are applied and manoeuvres of other vehicles needed to be considered. Interaction-aware motion models are more robust to predict collision, though they take more time and costs to compute all the trajectories in a time segment from large amount of data.

Figure 2.2 illustrates the difference between three types of motion models proposed by Lefèvre et al. (2014). Physics-based and maneuver-based motion models do not consider other vehicles manoeuvre. Interaction-aware motion models predict motion based on inter-vehicle dependency and are constrained by traffic rules. Interaction-aware motion models are addressed in this section because this thesis aims to identify or predict dangerous driving behaviour which is measured by inter-vehicle dependency (speed and manoeuvre of both cars).
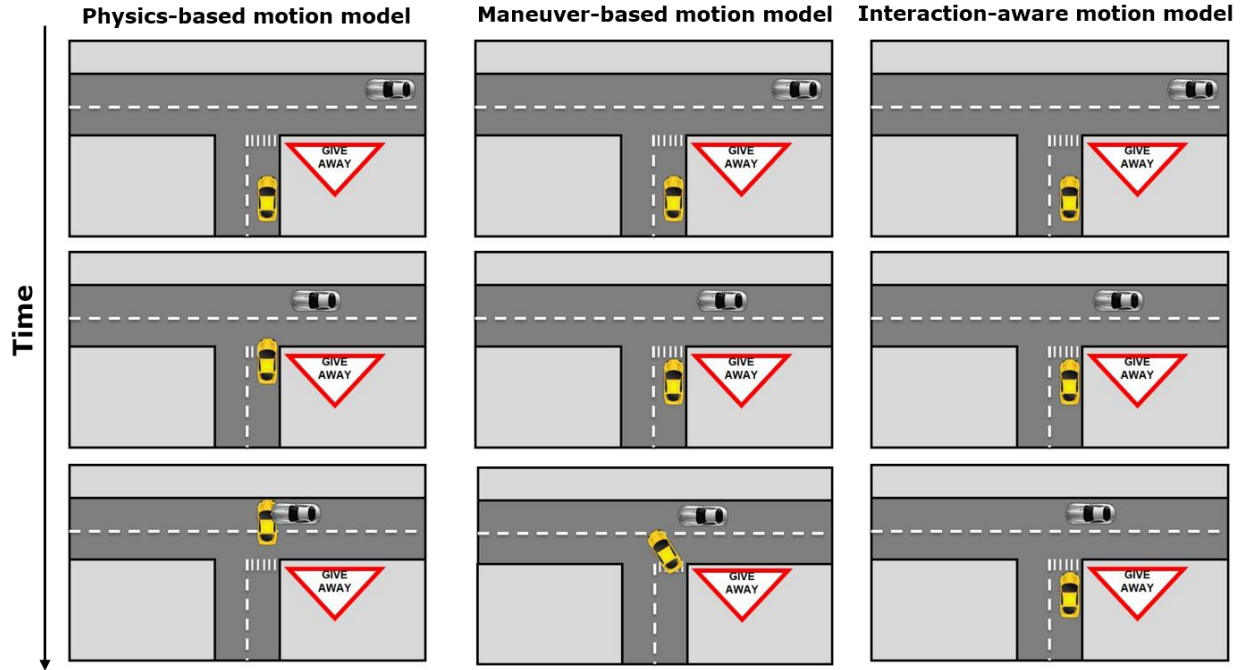
Figure 2.2 Motion prediction by three types of motion models (Lefèvre et al. 2014)

Most of the interaction-aware models are based on Dynamic Bayesian Network (DBN), because of their performance of handling missing data and representing the relationship between variables and outcomes (Lefèvre et al. 2014). DBN based models are widely used in literature for real-time probabilistic collision prediction. Worall et al. (2012) built a probabilistic motion model by using DBN with surrogate safety measure TTC. However, their model was not built to perform well under complex traffic scenarios and V2V communication was assumed. This limitation is overcome in the research of Gindele et al. (2015), where inter-vehicle communication was integrated in the car-following models for better perception of vehicle motion. They also used TTC for the risk assessment. But their model needed too many network level variables like road geometry, other cars states and traffic rules. To reduce the effect of variables, a static street model for motion planning was proposed by Kuhnt et al. (2015); although their model did not describe V2V communication efficiently. Bahram et al. (2016) stated that only two variables- road geometry and traffic rules can enhance the prediction time of manoeuvre perception, even without V2V communication. However, their approach lacked the ability to predict other vehicle's manoeuvres with abrupt acceleration or decelearation in complex road geometry.

2.1.2 Decision-making approaches for manoeuvre choice

Decision-making modules for manoeuvre planning are based on the modelling of traffic environment (Katrakazas et al. 2015). Different approaches are found in literature for the decision-making of autonomous vehicles. Multiple Criteria Decision Making (MCDM) approach was used by Furda and Vlacic (2011) for decision making when executing manoeuvres. Their approach needed several criterions like traffic rules, road boundary, safety distances and required accurate measurements from sensors and inter-vehicle communication. Kala and Warwick (2013) assumed in their approach that surrounding vehicles are non-autonomous, and the road has no lane. Different manoeuvres were predetermined and they were displayed according to the motion of other traffic participants. This decision-making approach was only for straight roads and showed a delay in the decision-making for some manoeuvres like overtaking or centring on the curves/bends.

Wei et al. (2014) used Prediction and Cost-function Based (PCB) approach for decision making. They generated multiple trajectories and chose the best one based on safety, comfort and fuel consumption. They validated the approach in both simulation environment and on-road, considering motion of other traffic participants and time delay. The approach showed 90% reduction on computational costs, although only single lane behaviours were considered.

Markov Decision Processes (MDP) were employed by White and White (1989) for selecting the best manoeuvre while driving. MDPs choose the best manoeuvre from a set of predetermined manoeuvres performing under uncertainty by maximizing weights. But MDPs assume all the states are fully observable. This limitation can be overcome by the approach of Partially Observable Markov Decision Processes (POMDP) which assumes a state of vehicle is unknown (Ong et al. 2010). Hubmann et al. (2017) used POMDP to select manoeuvre from unknown manoeuvre intentions. Their approach did not rely on V2V communications and operated on a continuous state space. However, their approach is not efficient for more than three traffic participants at a time. Mixed Observability MDP or MOMDP is also used if some states are unknown. MOMDP was used in predicting the intention of pedestrians while interacting with AVs by Bandyopadhyay et al. (2013). This approach assumed that the position and velocity of each pedestrian are known and their intentions remain constant. Brechtel et al. (2014) used a continuous POMDP approach and assumed that majority of the states are known. They simulated the model in merging cases where vision is obstructed due to illegal parking. However, their approach needed a lot of samples and a lot of options were available for decision-making.

Interactions between vehicles while making decision for the best manoeuvre can be introduced by Game Theory. Aoude et al. (2010) used Game Theory to build a threat assessment model in RRT planner for intersection by choosing the best path to avoid collision. Time-to-Collision was used as the safety indicator and speed limit of 0.5 m/s was selected. This approach also assumed no uncertainty in obstacle's motion.

Martin (2013) also used the same approach of Game Theory to predict the motion of other traffic participants in highways. Position, speed and acceleration of vehicles were taken as model criterions and the best manoeuvre was decided from selected manoeuvres of driving straight and changing lane. However, the approach assumed the lane to be straight and up to four vehicles could be simulated at a time.

Most of the above approaches of motion planning and decision-making have the limitations in handling obstacles and perception ability. Majority of the approaches assume no uncertainty. Also, interaction among vehicles is often assumed. Therefore, new approaches are needed for better perception capability and obstacles handling to improve risk assessment of AVs.

2.2   Real-time collision prediction approaches

As stated in 2.1, research in improving the safety of autonomous vehicles has got an immense attention in the present world. The success of AV depends on the acceptance of people and for that reason, their safety needs to be ensured or properly addressed. With constant improvement in Intelligent Transport Systems (ITS), risk assessment of AVs by real-time collision prediction has become a hot topic for the researchers. Real-time collision prediction is based on the concept of estimating the probability of collision during a short-time prediction state (Abdel-Aty and Pande 2005). Pande et al. (2011) stated that collision risk prediction can be done by comparing the traffic measurements like speed, traffic flow, occupancy etc. on a specific road segment of a time segment just before a collision with normal situation's traffic measurements of the same segment and time. This collision risk prediction is called Network Level Collision Prediction (NLCP), where future collision-prone situations are identified by studying real-time traffic data and road environment (Hossain 2011). Number of collision risk assessment approaches are found in the literature based on NLCP. But to understand the full safety challenge for automated vehicles, vehicle level collision

risk prediction (safe speed, acceleration, braking of the individual vehicle) is also required. Agamennoni et al. (2012) stated that there is a lack of research that addresses the collision risk based on vehicle level for real-time risk assessment of autonomous vehicles. Lefèvre et al. (2012) proposed a risk assessment model based on Dynamic Bayesian Network (DBN) which compares driving behaviour at safe and dangerous situations and predicts the collision by studying the vehicles, which were going to be in danger or in potential collision. Their model did not include trajectory of other vehicles and hence indicated that the collision prediction model can be enhanced by studying dangerous driving scenarios without fully depending on trajectory studies of all the vehicles in a road segment. Recently, Katrakazas (2017) developed an advanced DBN based collision risk model for AVs to integrate both network level collision risk and vehicle level collision risk.

Traffic conditions and road geometry affect types of collision enormously (Golob and Recker 2004); hence separate collision prediction models are required for different types of collisions (Qu et al. 2013). Typically collisions can be classified into four categories named rear-end, side sweep, head-on and right angle collisions (Mohamed et al. 2017). Among these four types, rear-end collisions mostly occur in roads. In USA over 2.5 million rear-end collisions are reported every year which is one-third of all the collisions (Singh 2003). In Japan, 35% of the collisions at intersections are due to rear-end collisions (Wang et al. 2003). It is obvious from literature that 60%-90% rear-end collisions could be avoided if drivers were warned 0.5-1 second before the collision occurred (Meinel 1998).

According to Lefèvre et al. (2014), collision prediction is of two types- binary collision prediction and probabilistic collision prediction. Several risk indicators or safety measures are used in predicting collision. Real-time collision prediction can be done by two methods named statistical methods and artificial intelligence methods (Hossain and Muromachi 2012). Traffic data from collision-prone situations and safe situations are compared to find out the indicators or predictors of collision in statistical methods. Abdel-Aty et al. (2004) used a method called matched-case control to remove the effects of environment on collision occurence probability. Individual vehicle level study at normal and collision-prone situations is the core objective of this method. Later this method was used with logistic regression to predict collision occurance by different authors (Abdel-Aty et al. 2005, Lee et al. 2006, Zheng et al. 2010, Pande and Abdel-Aty 2007). Aggregate log linear model (Lee et al. 2003) and Bayesian statistics (Oh et al. 2001) are also used to predict collision after statistically comparing traffic data of safe and collision-prone situation. However, statistical methods are not highly effiicient in predicting collision because it neglects most of the traffic flow variables like speed, traffic flow, occupancy (Hossain and Muromachi 2012). Therefore, artificial intelligence methods or machine learning classification models for real-time collision prediction are highly practised in recent time.

Artificial intelligence methods or machine learning algorithms are immensely popular in big data analysis now a days. They can learn from thousands of data and aim at identifying a vigorous description of a dataset given a limited sample (Herbrich 2001). They are capable of handling missing data and predicting collision based on learning from a large dataset in a short time. Classification or pattern recognition is one type of supervised machine learning technique which predicts the outcome of a given sample (Murphy 2012). Classification can be defined as the task of approximating a mapping function (f) from input variables (x) to discrete output variables (y) (Asiri 2018). Different machine learning classifiers are used in previous literature to predict collision risk in real-time.

Researchers tried to build different risk assessment models based on several machine learning classification techniques for predicting real-time rear-end collision. Oh et al. (2001) built a crash likelihood prediction model based on Bayesian classification, where they considered 5-minute standrad deviation of speed before crashes as the variable or crash predictor. Neural Network (NN) have been used by different researchers

for real-time collision prediction. Probabilistic Neural Network (PNN) classification technique was used by Abdel-Aty and Pande (2005) for identifying crashes, taking speed variations as the predictor. Their model could predict 70% of crashes. In continuing work, they integrated multi-layer perceptron (MLP) and normalized radial basis function (NRBF) with neural network to predict rear-end collisions (Pande and Abdel-Aty 2006). However, their model produced a significant number of false alarm (false collision prediction). Limitations of Neural Network approaches are over-fitting (Yu and Abdel-Aty 2013) and 'black-box' effect, which affects model transferability and interpretation of results (Sargent 2001). k-Nearest Neighbor (kNN) was introduced in different works to reduce this 'black-box' effect (Lv et al. 2009, Lin et al. 2015, Katrakazas 2017). A brief description of k-NN is given in section 4.4.4.

Support Vector Machine (SVM) has been used in recent times by different researchers in real-time safety assessment to predict rear-end collision. SVM classifier gave better accuracy in predicting crash frequency than traditional binomial models in the research of Li et al. (2008). SVM also performed better in crash injury severity analysis than Ordered Probit (OP) model in the work of Li et al. (2012). Mima et al. (2009) used SVM in building a warning system for rear-end collision. Aoude et al. (2012) used SVM and hidden Markov model to predict dangerous driving behaviour at intersection and validated the results on naturalistic data. Three SVM models, based on radial basis, sigmoid and polynomial kernel functions have given great accuracy (88.5%, 78.2% and 79.5%) in real-time freeway sideswipe crash prediction by Qu et al. (2013). SVM is also used successfully in other studies for real-time collision prediction (Yu et al. 2013, Wang et al. 2013). SVM is less susceptible to over-fitting problem when classifying large driving dataset. Mechanism of SVM is described in section 4.4.3.

Relevance Vector Machines (RVM) is another technique which performs like SVM (Bishop 2006). It is used in very few approaches of collision prediction. Katrakazas (2017) used RVM with SVM in network level collision prediction model. In their recent work, Katrakazas et al. (2019) used Random Forest and Neural Network classifiers, integrated with imbalanced learning for real-time collision prediction from raw speed time-series data from a driving simulator. Their study has showed that Random Forest performs well in predicting collision-prone situation and also imbalanced learning approaches can improve the classification performance as much as 40%.

Bayesian Networks (Hossain and Muromachi 2012) and Dynamic Bayesian Network (DBN) (Katrakazas 2017, Sun and Sun 2015) have been used in recent approaches of real-time collision prediction extensively because of their ability to represent the dependency between predictors and dependent variables. However, Bayesian approaches need a large dataset for training safe and collision-prone situations and might be ineffective in low number of collision-prone situations.

2.3   Surrogate safety measures

Surrogate safety measures are widely used to assess the safety of a road before any collision occurs (Ariza 2011). Surrogate measures are not reliable on crash data and these are used to analyze danger or collision risk based on crash records (Ceunynck 2017). Traffic accident events are rare. As surrogate measures don't need any crash to happen, they are widely accepted in road safety and collision prediction research. Many studies have shown surrogate safety measures are more proactive, accurate and time-efficient than crash based analyses (Hydén 1987). As crash data based analyses lack of real-crash data and don't contain behavioural factors, which can lead to collision, surrogate safety measures have become a lot popular in risk assessment (Laureshyn et al. 2017).

Perkins and Harris (1968) first introduced the idea of surrogate safety measures. With the immense improvement in sensor techniques and driver assistance systems, these indicators have been applied with success in recent times because data collection has become more efficient (Laureshyn et al. 2010, Saunier

et al. 2010, Tarko et al. 2017). Many indicators are used as Surrogate safety measures like velocity, traffic volume, gaps, delay etc. Typical time-based surrogate indicators are widely used to describe safety status from vehicle level data. In this section, various time-based surrogate safety indicators "Time-To-X" (or TTX) is briefly described, where X corresponds to an event in the course toward the collision (Lefèvre et al. 2014).

2.3.1 Time-Headway (THW)

Time-to-Headway is one of the most traditional time-based surrogate safety indicators, which is defined as the time difference between two consecutive vehicles in the same lane (Yang 2012). The equation can be described as below:

$$THW = t_i - t_{i-1} \tag{2.1}$$

Where t_i and t_{i-1} show the time of the following vehicle and lead vehicle pass in a lane respectively in seconds.

If the difference between the position of two vehicles (Δd) and the speed of the following vehicle (V_i) is known, THW can be derived as:

$$THW = \frac{\Delta d}{V_i} \tag{2.2}$$

In previous studies, different values of THW have been taken as a critical value under which the vehicles would have a collision risk. Some studies show that if the time gap between two vehicles is less than 1 second, the situation is unsafe. THW between 1.1 and 1.7 seconds is defined as 'comfortable' gap by Ohta (1993), while less than 0.6 sec was defined as an area of danger. To avoid rear-end collisions minimum headway of 2 seconds should be maintained (Evans 1991, Michael et al. 2000). Many European road administrations have recommended 2 seconds as a safe headway (Vogel 2003).

2.3.2 Time-to-Accident (TA)

According to Hydén (1987), Time to accident (TA) is defined as the time between when an evasive action was taken (like harsh brake, deceleration) and when a collision would have occurred if the two road users would continue with same speed and direction. If the following vehicle speed is Vi and the gap between the two vehicles is Δd, minimum TA for the following vehicle coming to stop is defined as (assuming the vehicle can successfully stop at collision point):

$$TA = \frac{\Delta d}{Vi} \tag{2.3}$$

Different threshold values are considered for critical TA to determine the seriousness of conflicts. TA value of 1.5 seconds was taken by Hydén (1977) to distinguish between serious and slight conflict. However, Shbeeb (2000) found that in urban areas 1.5 seconds limit is regarded fine but not in rural areas because of high speeds. Hydén (1987) later took 0.5 seconds as a limit of TA and described the severity level of collision compared to 0.5 seconds as well as Archer (2005) defined the same.

2.3.3 Post-Encroachment Time (PET)

Post-Encroachment Time (PET) is defined as the time between the first road user leaving the common spatial zone and the second road user arriving at it (Allen et al. 1978). As shown in Figure 2.3, PET is illustrated as the time from the first vehicle leaving the conflict spot to the second one reaching at the spot, measured from the rear-end of the first vehicle to the front-end of the second (Songchitruksa and Tarko 2006).
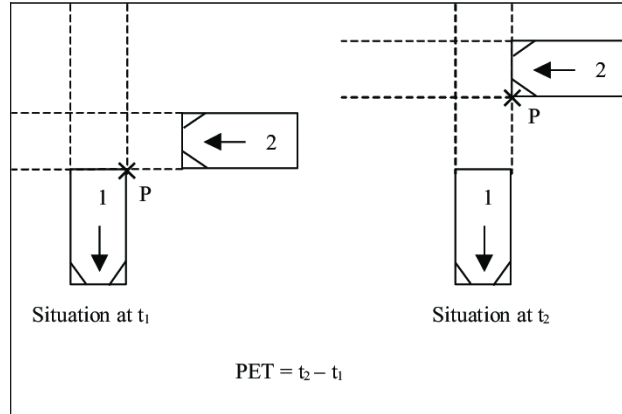
Figure 2.3 Illustration of Post-Encroachment Time (van der Horst 1990)

Threshold value of PET is normally taken as 1 or 1.5 seconds (Archer 2005). Extraction of PET is easier because speed and distance are not measured for the calculation of PET (Yang 2012). PET is a good surrogate measure, because driving interaction with other users is taken into account (Tarko and Songchitruksa 2005).

The drawback of PET is, it cannot be applied to a large area, because it is based on data collected from a fixed spatial zone (Abbas and Khan 2007).

2.3.4   Time-to-Collision (TTC)

TTC is one of the most popular indicators in Collision Avoidance system (van der Horst and Hogema 1993). It is defined as the time required for two vehicles to collide, if they continue at their present speed and on the same path (Hayward 1972). If i is the following vehicle and $(i-1)$ is the lead vehicle, TTC is calculated by the gap of the two vehicles divided by the speed difference of following vehicle and lead vehicle. The equation for deriving TTC (Yang 2012) is given below:

$$TTC_i(t) = \frac{X_{i-1}(t) - X_i(t) - L_i}{V_i(t) - V_{i-1}(t)} \qquad \forall V_i(t) > V_{i-1}(t) \qquad (2.4)$$

Where X denotes the position and V denotes speed. L_i is the length of the following vehicle.
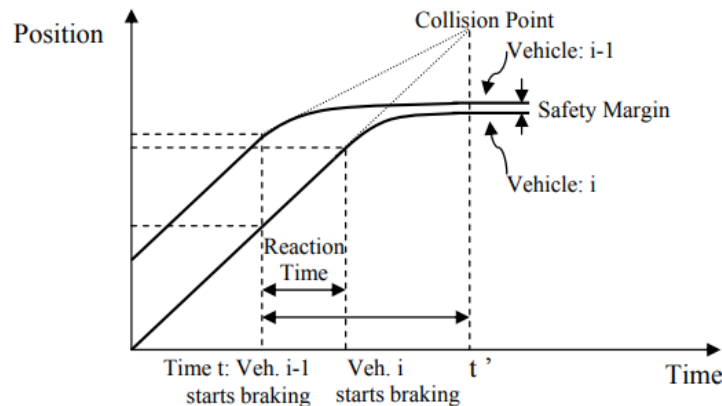
Figure 2.4 Vehicle trajectories and TTC (Yang 2012)

TTC is measured by predicting the future motion of two road users, which are on a collision course (Ceunynck 2017). In calculating traditional TTC, only vehicle speed is used and acceleration is not

considered (van der Horst 1990). As a result, only a positive speed difference between the following vehicle and lead vehicle must be taken to calculate the TTC. If the speed difference is zero, TTC will be infinity. So, this case is also omitted. A threshold TTC value is to be defined for differentiating dangerous and safe situation.

TTC is more accurate and popular surrogate safety indicator than other indicators like PET, THW or TA. THW only shows the potential danger but does not show the actual situation of an occurrence like low TTC value does (Vogel 2003). As described in 2.3.3, PET indicator has less correlation with conflict because of scatter behaviour (Lord 1996). TTC gives more information about dangerous situations than PET because PET cannot be applicable to all interactions (Archer 2005). TTC has been taken as an important factor by various automobiles to design collision avoidance systems (Riener and Ferscha 2009, Gettman and Head 2003).

TTC value defines how close the following vehicle is approaching to the lead vehicle. Low value of TTC means the distance is decreasing and braking or deceleration is required. However, TTC value cannot define the severity of conflicts. Two same value of TTC defines same severity although speeds of vehicles can be different (Archer and Young 2010). Previous studies have provided various values for the threshold of TTC. Some studies have shown 4 seconds as a safe situation (Farber 1991, Horst 1991), while some others have taken 1.5 seconds as the minimum value for avoiding Collision (Hydén 1987, Lord 1996). However, there is no fixed value to define minimum TTC in literature.

TTC has been used to derive other time-based indicators like Time-to-Zebra (Várhelyi 1998), Time-to-Line crossing (van Winsum et al. 2000), Inverse TTC (Balas and Balas 2006, Kiefer et al. 2005), Time Exposed TTC (TET) and Time Integrated TTC (TIT) (Minderhoud and Bovy 2001). However, they are not widely used indicators in literature like TTC.

2.3.5   Deceleration family

Deceleration family has different indicators which are rarely used in literature to predict dangerous situation. Initial Deceleration Rate is mostly used among these indicators. It is the magnitude of deceleration when the driver starts evasive braking (Ceunynck 2017). Higher deceleration rate indicates high probability of collision occurrence. In Figure 2.5, the Initial Deceleration Rate is measured as the second derivative of Curve B at time t2 (Gettman and Head 2003).

Another indicator is Maximum Deceleration. It is the maximum value of deceleration during a collision course (Gettman and Head 2003). Deceleration-to-Safety Time is defined as the minimum required deceleration to avoid collision (Hupfer 1997). These indicators are discussed in the literature although no validation is provided with crash data.
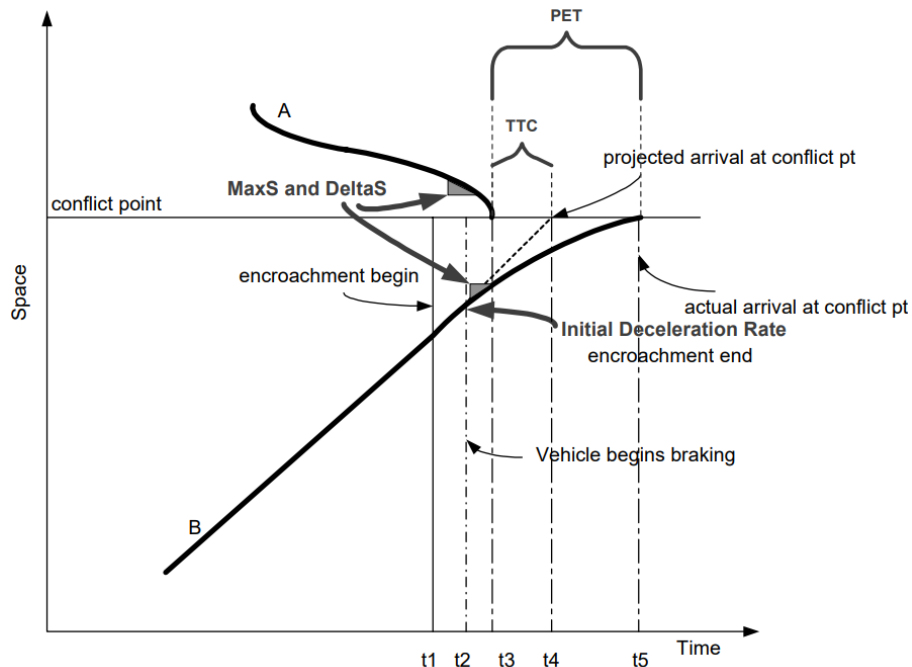
Figure 2.5 Different surrogate measures defined on conflict diagram (Gettman and Head 2003)

2.4   Safety assessment using naturalistic driving data

Over the past years, safety assessment has been researched extensively by studying safety-related events (Shankar et al. 2008, Wu and Jovanis 2012, Wu et al. 2014). Human errors are mostly responsible for traffic accidents; hence, driving behavior, driver characteristics and collision involvement has been studied in literature using safety-related events (Evans and Wasielewski 1983, Waggenaar and Reason 1990, Verschuur and Hurts 2008, Wu et al. 2014). Early studies of safety-related events were based on crash data. The need for surrogate events related studies arises, because crash data samples are rare and surrogate events include both crash and near-crash samples (Wu and Jovanis 2012, Laureshyn et al. 2017, Ceunynck 2017). Previously, safety-related events studies were done either by collecting field data from intersection and segments or by interviewing drivers (Wu et al. 2014). Video-tape and street cameras were used for data collection at intersections or road segments (Hydén 1987, Chin and Quek 1997). However, these techniques of data collection did not include driving behaviour and driver distraction. The outcome from these data could argue that safety-related events occur more frequently than normal driving scenarios (Hanowski et al. 2005); which is completely opposite in driving scenarios. Also, risk assessment studies need to focus on the events caused by driving error (Wåhlberg 2003). For better safety assessment, Verschuur and Hurts (2008) have proposed to integrate driving behaviour and driver characteristics studies with collision involvement.

For vehicle level collision prediction, driving exposure or driver behavior study is important, because sometimes high speed or acceleration does not lead to a collision. For better prediction of collision-prone situations, this dangerous driving behaviour data is needed to be included in the model. Naturalistic Driving Study (NDS) has become a big advancement in data collection techniques for safety assessment which helps by providing data to understand driver behaviour (Muronga and Venter 2014). NDS provides precise observation and measurements of safety-related events (Dingus et al. 2006). According to UDRIVE (European Naturalistic Driving Studies), NDS can be defined as *"A study undertaken to provide insight into driver behaviour during every day trips by recording details of the driver, the vehicle and the*

surroundings through unobtrusive data gathering equipment and without experimental control". Under this approach, behaviour of road users and drivers is observed unobtrusively for a period of time. The vehicles are equipped with devices (cameras and sensors) for continuous monitoring of driving behaviour as well as vehicle dynamics like speed, position, acceleration, yaw rate etc. NDS help to observe the interrelationship between road users, vehicle and vehicle surroundings which are the three main physical characteristics of the road transport system. NDS can be used to analyze driving behaviour as well as in collision prediction model to advise drivers for taking necessary actions preceding crashes or near-crash events (Venter 2014).

The 100-Car Naturalistic Driving Study (from 2004-2005) by Virginia Tech Transportation Institute is the first large-scale NDS in USA to collect data about crashes and near-crashes for using in collision risk prediction model (Dingus et al. 2006). The study included 102 primary drivers in northern Virginia and driving data were collected continuously for 12 months. The data acquisition system included five camera views (forward, driver face, over the shoulder, left and right mirror), GPS, speedometer, three-dimension accelerometer, and radar, etc. This naturalistic driving dataset contained about 2 million vehicle miles and 43000 hours of data (Neale et al. 2005). The results of the study provided information from 82 crashes and 761 near-crashes and 8295 incidents (less severe near-crashes). But due to the initialization of system and camera failure, some instances of crashes were missed in this study. 100-Car Naturalistic Driving Study dataset has been used by different researchers for risk assessment in road safety research. Guo et al. (2010) studied the association between crashes and near-crashes samples. Later they used this dataset to investigate risk factors associated with individual driving risk and built a risk prediction model by using logistic regression (Guo and Fang 2013). However, their study was limited by a sample size of 102 drivers. Klauer et al. (2006) utilized reduced samples of crashes, near-crashes and incidents data to identify driver's inattention. Wu et al. (2014) used single vehicle naturalistic driving data from the 100-Car NDS to find out interrelationship among driving behaviour, traffic-safety related events and collision involvement. Recently, Xiong et al. (2018) used this dataset to predict driving risk based on Markov Chain model.

The SHRP2 study (from 2011-2014) is another example of using NDS for safety assessment in USA (Gordon et al. 2013). Collision prediction in this study was done by statistical analysis of near-crashes and actual crashes based on surrogate safety measures. This study contained naturalistic driving data from 3000 volunteer drivers. Later different researchers used this SHRP2 dataset for safety assessment (Seshadri et al. 2015, Paone et al. 2015, Seaman et al. 2016, Hallmark et al. 2015, Wang and Zhou 2018).

The Australian 400 Car-NDS aimed at identifying driving behaviour at the time of normal and safety critical situations (Regan et al. 2013). The study has noted that NDS could affect driving behaviour because of continuous monitoring and therefore, very large data samples are required to yield sufficient outcomes.

In recent years, safety assessment using naturalistic driving data has become popular in road safety research (LeBlanc 2006, Gordon et al. 2013, Klauer et al. 2006). Driver's exposure to distraction was studied with unobtrusive video data collected from 70 drivers over one-week time period by Stutts et al. (2005). Hanowski et al. (2005) used naturalistic data from truck drivers to find out driver-distraction related collision-prone events. Shankar et al. (2008) proposed driver-based analysis of naturalistic driving data for better understanding of NDS paradigms. Some studies empirically tested these paradigms (Jovanis et al. 2012). Bender et al. (2015) utilized the same naturalistic dataset used in this thesis for predicting driver intention in intersection. In recent times, Molnar et al. (2018) used NDS to understand driving patterns among different age groups; although their study contained only 108 data samples. Li et al. (2018) validated a crash probability estimation model based on driver hazard perception ability on a naturalistic dataset containing two-month Collision Mitigation Braking System (CMBS) equipped vehicle data. However, different drivers took part in their NDS test, therefore individual driving characteristics was overlooked.

2.5 Identification of research gap from literature review

It is clear from the review of the literature that human driving error arising from misperception and wrong decision-making leads to most of the collisions. Autonomous vehicles are considered safer for reducing the part of human error. Therefore, accurate risk assessment of AVs is a prerequisite for preventing collisions. Current planning modules of AVs emphasize decision making which is a key to navigate the vehicle towards a safe manoeuvre and reach the destination bound by traffic rules. Through manoeuvre planning, AVs need to consider the behaviour of surroundings and take the best decision in navigation. Although current AV systems are considered to finding the safest and best navigation by detecting obstacles, collisions still occur. From the literature it is obvious that the most important challenge in collision risk assessment by AVs is handling of obstacles and perception and decision-making ability. Most of the approaches for risk assessment consider the vehicle as an independent entity and assume the presence of interaction among traffic participants. These approaches are based on traffic network related data and collision prediction is done by predicting the trajectories of vehicles. However, these approaches need a high amount of time and computational cost for checking all possible trajectories. Lefèvre et al. (2012) stated that the collision prediction can be enhanced by studying the scenarios of dangerous driving, without fully depending on trajectory studies of all vehicles in a road segment.

Most of the existing collision prediction models in AVs are based on road network level. Network Level Collision Prediction (NLCP) approaches are bound to road geometry and traffic rules, which cannot fully describe the underlying safety level regarding AVs. Moreover, the existing approaches face difficulty in collision prediction because of the high imbalance between 'safe' and 'collision-prone' situations. Hence, vehicle-level prediction approaches are needed for collision prediction in real time. Vehicle level collision prediction requires a large number of dangerous driving behavior samples based on vehicle dynamics such as speed, position, acceleration etc. Naturalistic driving data from NDS have been immensely used in collision risk assessment due to the inclusion of driver characteristic and driving behaviour data as well as precise observation of safety-related events. However, most of the safety assessment approaches using NDS face the problem of small data samples or monitoring of different groups of drivers which overlook to predict collision from individual dangerous driving behaviour.

Existing vehicle level collision prediction approaches try to predict collision either by identifying safe and collision-prone situations from statistical analysis or by machine learning classification techniques. Machine learning approaches are more popular in recent times for real-time collision prediction because of their effective handling of missing data. However, easy interpretation of results is sometimes hindered due to 'black-box' effects and some approaches suffer from overfitting and high imbalance between samples. Hence, a robust vehicle-level approach easily integrated into behaviour planning of AVs, is needed in order to predict dangerous situations with regards to the context from individual vehicle data.

3   Data Collection and Visualization

The dataset used in this thesis is derived from a vehicle driving around single-lane urban streets around the Australian Centre for Field Robotics in Sydney (http://its.acfr.usyd.edu.au/datasets/naturalistic-intersection-driving-dataset/). The dataset contains the position of the vehicle, speed and heading. Data were collected by a system fusing GPS and dead reckoning information from gyroscopes and odometry at a resolution of 10 Hertz. Position of the car was recorded at every 10 Hz by using global navigation satellite system (GNSS). Inertial data were collected by a strap-down inertial measurement unit (IMU) and speed data were recorded by wheel encoders of the car (Bender et al. 2015). The road feature of interest was an uncontrolled three-way T-intersection (shown in Figure 3.1). Three different drivers executed the driving manoeuvres by driving straight across the top of the intersection and turning into and out of the intersection both left and right for the purpose of predicting driving intent (Bender et al. 2015). They were instructed to drive through the T- Intersection approximately 10 passes in six possible manoeuvres (west-east, west-south, east-west, east-south, south-west and south-east). Trajectories in each manoeuvre are found out from heading data. Speed limit of the study area is 40 km/hour (11.11 m/s).
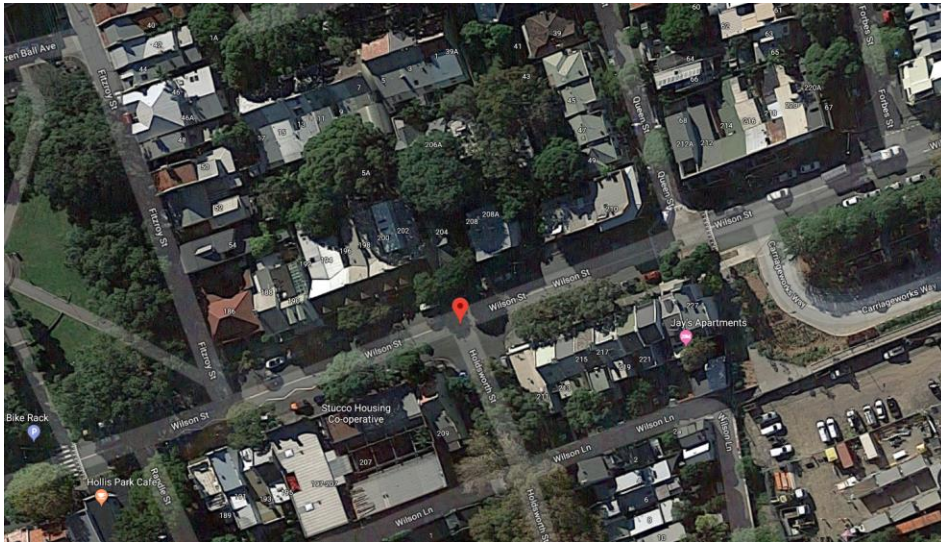
Figure 3.1 T- Intersection of the experiment area in google map

Figure 3.2 Collecting naturalistic driving data (http://its.acfr.usyd.edu.au/collecting-naturalistic-driving-data/)

The vehicle equipped with sensors used in this experiment is a mid-size Sedan (shown in Figure 3.2). Length of the vehicle is presumed to be 4.8 meters in the calculation of TTC, as the standard length of Sedan size. Position of the vehicle in each trajectory was extracted from latitude-longitude data. Trajectories of the vehicle for three different drivers are visualized in OpenStreetMap using position data in figure 3.3, 3.4 and 3.5 respectively. The T-intersection is visualized by the circle area.
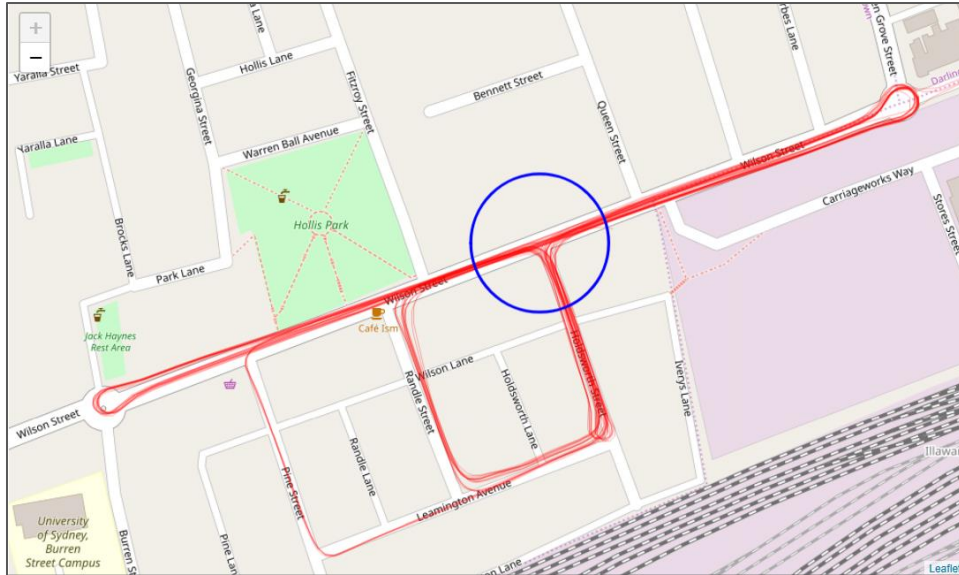
Figure 3.3 Collected trajectories of driver 1 in OpenStreetMap in Python environment
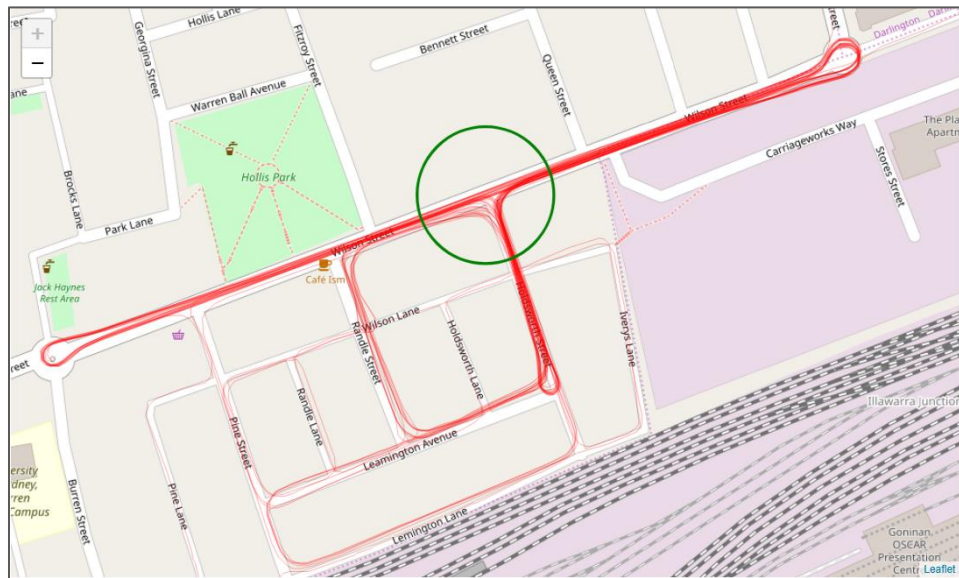
Figure 3.4 Collected trajectories of driver 2 in OpenStreetMap in Python environment
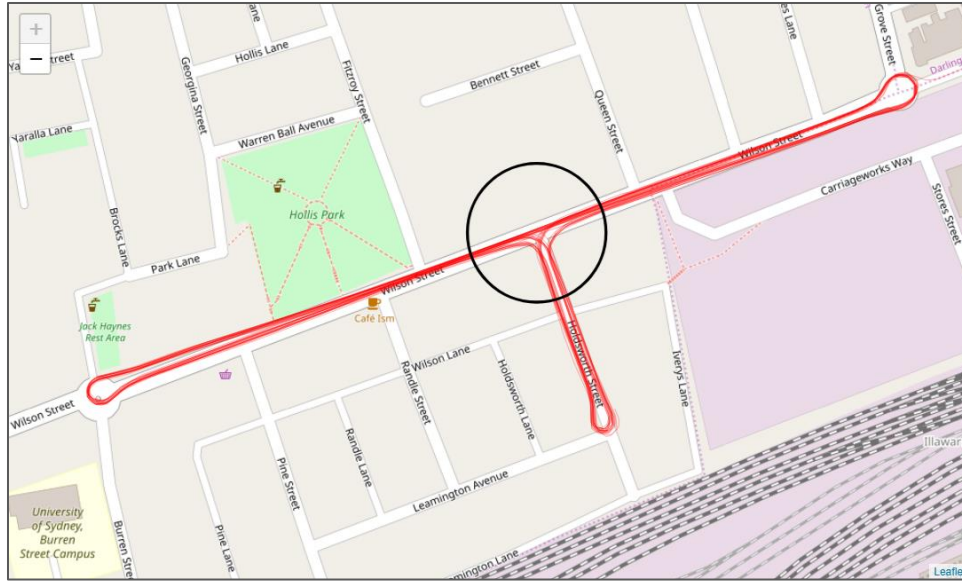
Figure 3.5 Collected trajectories of driver 3 in OpenStreetMap in Python environment

Table 3.1 shows number of trajectories in the circle area in six possible manoeuvres by three different drivers. In total 199 separate manoeuvres were found. Number of trajectories found for driver 1, driver 2 and driver 3 was 69, 71 and 59 respectively.

Manoeuvre	Driver 1	Driver 2	Driver 3	Total
East-west	14	11	10	35
East-south	10	14	9	33
West-east	13	11	10	34
West-south	10	11	10	31
South-east	11	14	10	35
South-west	11	10	10	31
Total	69	71	59	199

Table 3.1 Number of trajectories in bounding area

For every driver, speed data were analyzed so as to understand the speed distribution of the vehicle. Results are shown in Table 3.2.

21

Speed (m/s)	Driver 1	Driver 2	Driver 3
Total count	443318	602331	307834
Mean	6.974386	5.175807	7.621625
Std deviation	3.697586	2.965245	3.411286
Maximum value	14.973695	13.343713	13.397516
Minimum value	0	0	0

Table 3.2 Speed profile for three different drivers

As stated, all the three drivers were instructed to pass through the intersection approximately 10 times in all the six manoeuvres to get more interaction points among the vehicle trajectories, so as to extract sufficient car-following events. From Table 3.1, it is evident that the first two drivers passed through the intersection more than 10 times in some manoeuvres. The third driver passed through the intersection 10 times in each manoeuvre (except east-south = 9) and results in same number of trajectories in the six manoeuvres (except east-south). That is why, sensor data from the third driver are used in this thesis for generating car-following events to identify and classify dangerous driving behaviour. Speed variation of the third driver is shown in Figure 3.6.

Figure 3.6 Speed variation with time for driver 3

4   Methodology

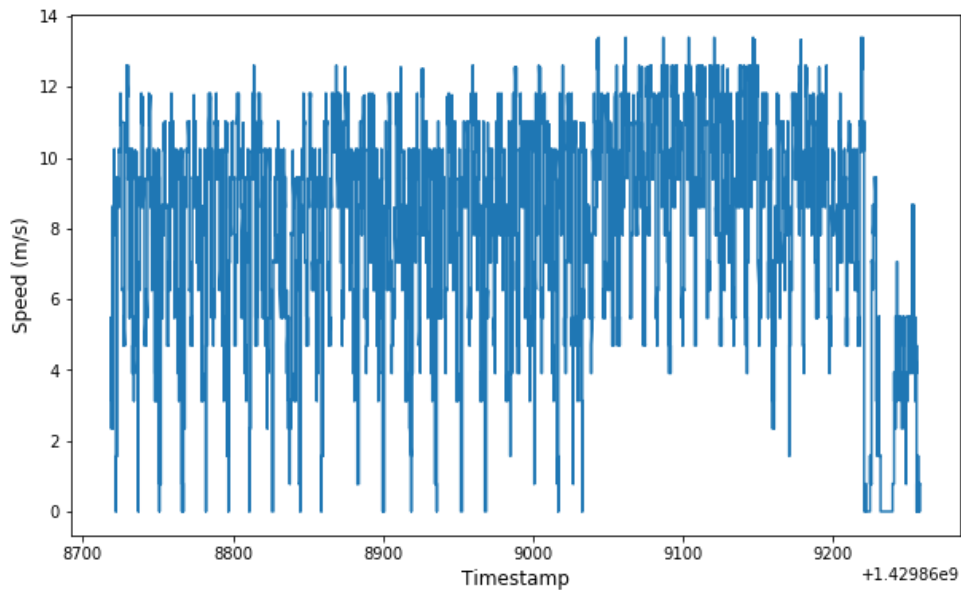## 4.1   Trajectory extraction in bounding area

To identify dangerous driving behaviour in a roadway based on vehicle level data, relative distance and speed of the vehicles are needed. High speed can be more collision prone in the intersection than in highways (Worrall et al. 2012). Intersections are rated as a highly collision-prone area of the road network due to the occurrence of multiple conflicting manoeuvres (Sobhani et al. 2013). In this thesis, the uncontrolled T-intersection is regarded as the area to identify dangerous situations, because an uncontrolled intersection is a complex scenario with high collision risk (Song et al. 2016).

A boundary region is defined around the intersection to get interaction points among trajectories for modelling driving behaviour. For each dataset, the influence region is taken as 35 meters at each side from center of the intersection. When identifying dangerous driving behaviour, vehicle trajectories in this circle are only considered. Trajectories are extracted based on entry and exit points from the bounding area. Rest of the data outside the area are discarded by using programming language python. Thus, trajectories are found for different manoeuvres in the bounding area, which are used to identify dangerous driving behaviour. Figure 4.1 visualizes the trajectories in the boundary area for the third driver.

Figure 4.1 Trajectories in the bounding area for driver 3

4.2   Concept of time shifted trajectory

To estimate time-based surrogate safety indicator TTC, car-following events are required. For identifying dangerous driving scenarios, we also need vehicles driving very closely. As it is not feasible to generate trajectory data for actual collisions, time shifting of trajectory method is used to find out collision probability (Ward et al. 2014a). In this method, time shifting is controlled to get the trajectory segments to pass each other closely or collide at some point. Each trajectory in a manoeuvre can be considered as a new vehicle trajectory and sufficient car-following events can be generated for the experiment. In this thesis, the first trajectory entering the bounding circle in each manoeuvre is controlled to pass closely or have rear-end collision with other trajectories in that manoeuvre and considered as lead vehicle. Shifting of the trajectories is controlled by maintaining an initial safe distance between the lead vehicle and the following

vehicle trajectory. In this thesis, initial minimum distance between the trajectory pairs is regarded as 4 meters, because radius of the circle of study area from center of the intersection is only 35 meters. Car-following events are then monitored in the bounding area for identifying driving behaviour from all pairs of the trajectories in each of the six manoeuvres. Car-following events are considered only when positive speed difference exists between the following and predecessor car; otherwise, collision never occurs (Yang 2012). So, for identifying and classifying driving behaviour in this thesis, negative speed difference cases are omitted in the car-following events in order to avoid high imbalance between safe and dangerous samples.

4.3 Estimation of Time-to-Collision (TTC)

Time-to-Collision (TTC) is calculated in the bounding area in each direction to identify rear-end collision. As the bounding area is not very large, TTC values can be artificially bounded at 40 seconds for better monitoring of car-following events (Balas and Balas 2006). Threshold value of TTC is taken as 1.5 seconds because average human reaction time is 1.5 seconds (Triggs and Harris 1982). If TTC goes below the threshold, the situation is identified as dangerous. TTC value greater than 1.5 seconds is termed as safe. The equation stated in section 2.3.4 is used for the calculation of TTC.

$$TTC_i(t) = \frac{X_{i-1}(t) - X_i(t) - L_i}{V_i(t) - V_{i-1}(t)} \qquad \forall V_i(t) > V_{i-1}(t) \qquad (4.1)$$

Where, X_{i-1} and V_{i-1} = Position and speed of the lead vehicle respectively

X_i and V_i = Position and speed of the following vehicles respectively

L_i = Length of the following vehicle = 4.8 meter

Distance between two points is calculated by Haversine formula which is commonly used to find the distance between two different latitude and longitudes. It is a widely used formula in Geographic Information System (https://www.igismap.com/haversine-formula-calculate-geographic-distance-earth/(2015). Original Haversine formula is given below:

$$\text{Haversine} \left(\frac{d}{R}\right) = \text{haversin} (lat2\text{-}lat1) + \cos(lat1) \cos (lat2) \text{ haversin} (lon2\text{-}lon1) \qquad (4.2)$$

Here, d is the distance between two points and R is the radius of earth = 6371000 meters; lat and lon represents latitude and longitude respectively for two points. This formula can be derived as follows and is used in this thesis to calculate distance between the position of lead vehicle and following vehicle.

$$a = \sin^2 \left(\frac{lat1 - lat2}{2}\right) + \cos (lat1) * \cos(lat2) * \sin^2 \left(\frac{lon1 - lon2}{2}\right) \qquad (4.3)$$

$$c = 2*a*\tan2 (\sqrt{a}, \sqrt{(1-a)}) \qquad (4.4)$$

$$d = R*c \qquad (4.5)$$

The final equation derived from Haversine formula to find the difference between points in excel is given below in equation 3.6:

$$\begin{aligned} d = [ACOS\{COS\big(RADIANS(90 - lat1) * COS(RADIANS(90 - lat2)\big) \\ + SIN\big(RADIANS(90 - lat1)\big) * SIN\big(RADIANS(90 - lat2)\big) \\ * COS\big(RADIANS(lon1 - lon2)\big)\}] * 6371000 \end{aligned} \qquad (4.6)$$

As discussed in 4.2, by controlling the time-shifting technique, in total 11,208 gap observations were counted in the car-following events. TTC is then calculated by using equation 4.1. From Table 4.1, it is observed that 35.96% observations were marked as dangerous, where TTC lies below the threshold value

of 1.5 seconds. The most collision prone trajectories were found in the south-west and east-west manoeuvres, where almost half of the observations have TTC below the threshold. In west-east manoeuvre, there is no event where TTC goes below the threshold.

Manoeuvres	Number of trajectories	Number of events when TTC<1.5s	Rate of dangerous situation (%)	Total gap observations in Car-following events
South-east	10	1067	37.82	2821
East-south	9	277	14.45	1917
South-west	10	1210	50.48	2397
West-south	10	515	30.09	1711
East-west	10	962	45.48	2115
West-east	10	0	0	247
Total	59	4031	35.96	11208

Table 4.1 TTC-based evaluation results for driver 3

4.4   Classification of driving behaviour

In this research, different machine learning algorithms are employed to predict dangerous driving behaviour. Collision risk increases with high speed at intersections (Worrall et al. 2012). In this thesis, real-collision data is not used for safety assessment. Trajectories are controlled by time shifting to have interaction (e.g. collision) among the vehicles. For this reason, there is a chance of getting TTC value lower than the threshold of 1.5 seconds in some manoeuvres even when the speed of the vehicle is low. To avoid this effect, speed difference between the following vehicle and lead vehicle is taken as the only predictor in classification. Heading is not considered as variable because the change of heading in this dataset contributes very little to separate the data prior to entering the intersection (Bender et al. 2015). Two classes (dangerous and safe speed) are to be predicted, hence; this classification problem will give binary output (1 or 0). If we assume a training dataset $X_{training} = (x_n, y_n)$ where $n = 1,…N$ and x_n is an independent variable and y_n is the output in 0 or 1; a binary classification will predict new data from these training data and give the response to the correct class (Katrakazas 2017). In this section, different machine learning classifiers are briefly described which are tested and compared to each other in this thesis to predict driving behaviour.

4.4.1   Naive Bayes

Naive Bayes is a simple probabilistic classification technique based on Bayes theorem with assuming no independence among the predictors. Bayes theorem for probabilistic classification is given in equation 2.5

$$p(y|x) = \frac{p(x|y) \times p(y)}{p(x)} \qquad (4.7)$$

Where $p(y|x)$ is the posterior probability of class y given predictor x;

$p(y)$ is the prior probability of class y;

$p(x|y)$ is defined as likelihood or probability of predictor given class;

and $p(x)$ is the prior probability of predictor x

Naive Bayes is very simple to implement (Stewart 1998) and quick to predict class. When classifying numerical variables, it assumes the variables have normal distribution. So, it normally performs well in classifying categorical input variables. Although it was used in literature for real-time collision prediction, Naive Bayes classifier is considered as a bad estimator and does not perform well with increasing sample

size (Stewart 1998). It has also a problem of zero probability and cannot predict when conditional probability is zero for a variable (Gahukar 2018). Smoothing technique is needed to solve this problem.

In this thesis, two types of Naive Bayes classifiers are applied to predict dangerous behaviour, named Gaussian Naive Bayes and Bernoulli Naive Bayes. Generally, both are used in binary classification. There is another Naive Bayes classifier named Multinomial Naive Bayes which is used for discrete variable counts.

4.4.2   Logistic Regression

Logistic regression is a common classifier for binary classification which is also tested to predict the dangerous or safe driving behaviour in this thesis. It predicts and describes the relationship between one or more independent variables and one dependent binary variable and gives the outcome in 0 or 1. As the algorithm of this classifier is derived from highly interpretable linear regression, logistic regression has high interpretability (https://www.facebook.com/pages/Statistics-Solutions). For binary classification by logistic regression, it is assumed that there are no high correlations among the predictors. The correlation coefficients among independent variables have to be less than 0.90 (Tabachnick and Fidell 2007). Unlike Naive Bayes classifier, independent variables don't need to be normally distributed.

4.4.3   Support Vector Machine

Support Vector Machine (SVM) classifier separates training data into categories and new data are predicted to belong to a category based on which side they fall (Gahukar 2018). Support vectors are data points that lie closest to the decision surface and are the most difficult to classify (Berwick 2003). SVM classifies those points by separating hyperplane. There are some tuning parameters in SVM classification technique like Kernel, Regularization, Gamma and Margin. Kernel-based SVM is very useful in nonlinear classification (Murty and Raghava 2016). Parameter Gamma depicts how far influence of single training example reaches; that means with low gamma value distant points from the separation line are taken in calculation and with high gamma only points close to the separation line are considered in calculation (Patel 2017). SVM classifier performs well in high dimensional space and uses a subset of training points in prediction. It is only used in binary classification (gives binary outcome 0 or 1). They have been used successfully in literature for collision prediction because of their flexibility and sparsity when support vectors are much smaller in number than the number of training data (Ward et al. 2014). However, there is a 'Black-box' effect on SVM and Support Vectors grows linearly with increasing size of training sets (Bishop 2006, Dreiseitl and Ohno-Machado 2002). Separating mechanism of SVM can be visualized by Figure 4.2. It is observed that H1 and H2 hyperplanes do not separate the data classes in a good margin. H3 is used as the SVM here because it separates the classes by a good margin that means distance of both classes from the hyperplane are not large.
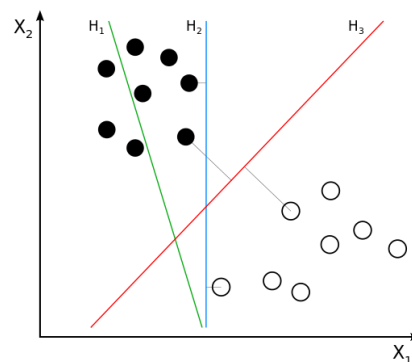

Figure 4.2 Support Vector Machine separating data by hyperplane (Gahukar 2018)

4.4.4 k-Nearest Neighbor

k-Nearest Neighbor (kNN) is a simple algorithm which can be used in both regression and classification problem. It is a popular classification technique in industrial research for it's simple interpretation in large training data and low calculation time (https://www.facebook.com/AnalyticsVidhya/ 2018). kNN classifies data to the class which is most common among its nearest neighbors k (k is the number of classes predefined). It is a lazy algorithm because it does not learn the model or make the generalization of data. It classifies the object based on feature similarity or input variables (Gahukar 2018). That's why it is not a very popular algorithm for predicting collision in literature.

4.4.5 Decision Tree

Decision Tree is a widely used classification technique which can be used to handle both numerical and categorical variables. It splits the data sample into two or more homogeneous sets based on the most significant differentiators or predictors in the input variables (Gahukar 2018). The Decision Tree algorithm chooses the predictor of highest accuracy by binary splitting all features and repeats the process until all the data samples in the leaves or sub-populations are trained. Figure 4.3 shows the leaves or subpopulations which are split by the predictor. After training all the samples, binary results are obtained.
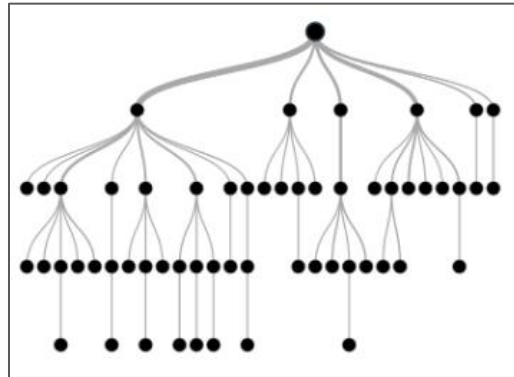

Figure 4.3 Decision Tree algorithm (https://www.facebook.com/AnalyticsVidhya/ 2016)

Decision Tree classifier is popular for its simplicity to understand and visualize. It also needs less data cleaning and is not influenced by outliers or missing values. Overfitting can be a drawback in classification by this technique. It can sometimes create a complex and unstable tree if any small variation in data occurs which will generate a completely different tree (Gahukar 2018).

4.4.6 Random Forest

Random Forest (RF) is a very useful algorithm for handling large data samples and can be used for both classification and regression. Bagging algorithms are used by RFs to create new training sets from the specific training set (Katrakazas et al. 2019). It creates decision trees on random samples, gets a prediction from each tree and then gives the best prediction by voting (Maheshwari 2019). As the output comes from the votes of all the trees, overfitting problem can be minimized (Gahukar 2018). It normally gives high accuracy in classification or prediction because a large number of trees give the final decision by voting. But the classification can be time-consuming for a large sample because of the large number of trees. Random Forest is successfully used in several research for real-time traffic safety assessment (Katrakazas 2017). Figure 4.4 illustrates the working algorithm of RF classifier where the output prediction comes from voting of different training sets.
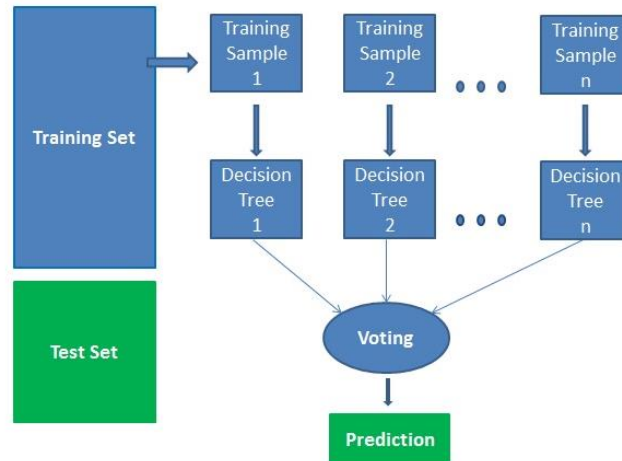
Figure 4.4 Random Forest classifier algorithm (Maheshwari 2019)

4.5   Performance evaluation of classifiers

As discussed in section 4.4, several classifiers are tested and compared to predict dangerous driving behaviour in this thesis. The output in the classification is binary (1 for dangerous and 0 for safe). Classification is done by using sci-kit learn in Python 3.7. Performance of the classifiers is evaluated by different metrices which are briefly discussed in this section. After evaluating performance, the best classifier is chosen and used in predicting driving behaviour.

4.5.1   Confusion matrix

Confusion matrix is used to assess the performance of a classification model. It is a table to describe the performance of a classifier on a test dataset for which true values are known. In a confusion matrix, predictions of each data are contrasted with the actual class to which they belonged, to ensure whether they are correctly classified or not. As dangerous driving behaviour is to be predicted, positive class represents 'dangerous' and negative class represents 'safe' behaviour.

Predicted label

Actual label		0	1
	0	True Negative (TN)	False Positive (FP)
	1	False Negative (FN)	True Positive (TP)

Table 4.2 Confusion matrix table

In this thesis, some of the popular matrices obtained from confusion matrix are used to compare the performance of classifiers. These matrices are defined below:

Accuracy: Accuracy is a very common and widely used measure. It is defined as the total number of values correctly classified (both safe and dangerous) divided by the total number of values classified.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \qquad (4.8)$$

High accuracy value does not always indicate high performance because it also indicates overfitting. As a result, accuracy is not taken as a widely accepted performance measure for the classifiers.

Precision: Precision is the proportion of the values that are actually dangerous. High precision means the classification model or classifier has high trustworthiness that means it produces less false alarms.

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (4.9)$$

Recall: Recall or sensitivity is the proportion of values known to be dangerous that test positive for it. High recall means the classifier can give timely warning.

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (4.10)$$

Specificity: It measures the proportion of 'safe' values that are correctly classified.

$$\text{Specificity} = \frac{TN}{TN+FP} \qquad (4.11)$$

False alarm rate: False alarm rate or false collision prediction rate is the percentage of 'safe' values that are classified as dangerous. It is calculated from specificity.

$$\text{False alarm rate} = 1\text{- Specificity} \qquad (4.12)$$

It can be also calculated from the following equation:

$$\text{False alarm rate} = \frac{FP}{TN+FP} \qquad (4.13)$$

f_1 score: It is the harmonic mean of precision and recall. f1-score is a more appropriate performance measure for large imbalanced dataset where difference between positive and negative cases is higher.

$$f_1 \text{ score} = \frac{2* \text{Precision}* \text{Recall}}{\text{Precision}+\text{Recall}} \qquad (4.14)$$

ROC curve: Another useful performance matrix is the Receiver Operating Characteristics (ROC) curve. It visually represents the trade-off between true positive rate and false positive rate (Asiri 2018). Accuracy of a classifier is measured by area under ROC curve (auc). A high value of area close to 1 defines the high accuracy of the classifier.

All the methods applied in this thesis from identifying to predicting dangerous driving behaviour by classification technique is illustrated chronologically by a flowchart in Figure 4.5.
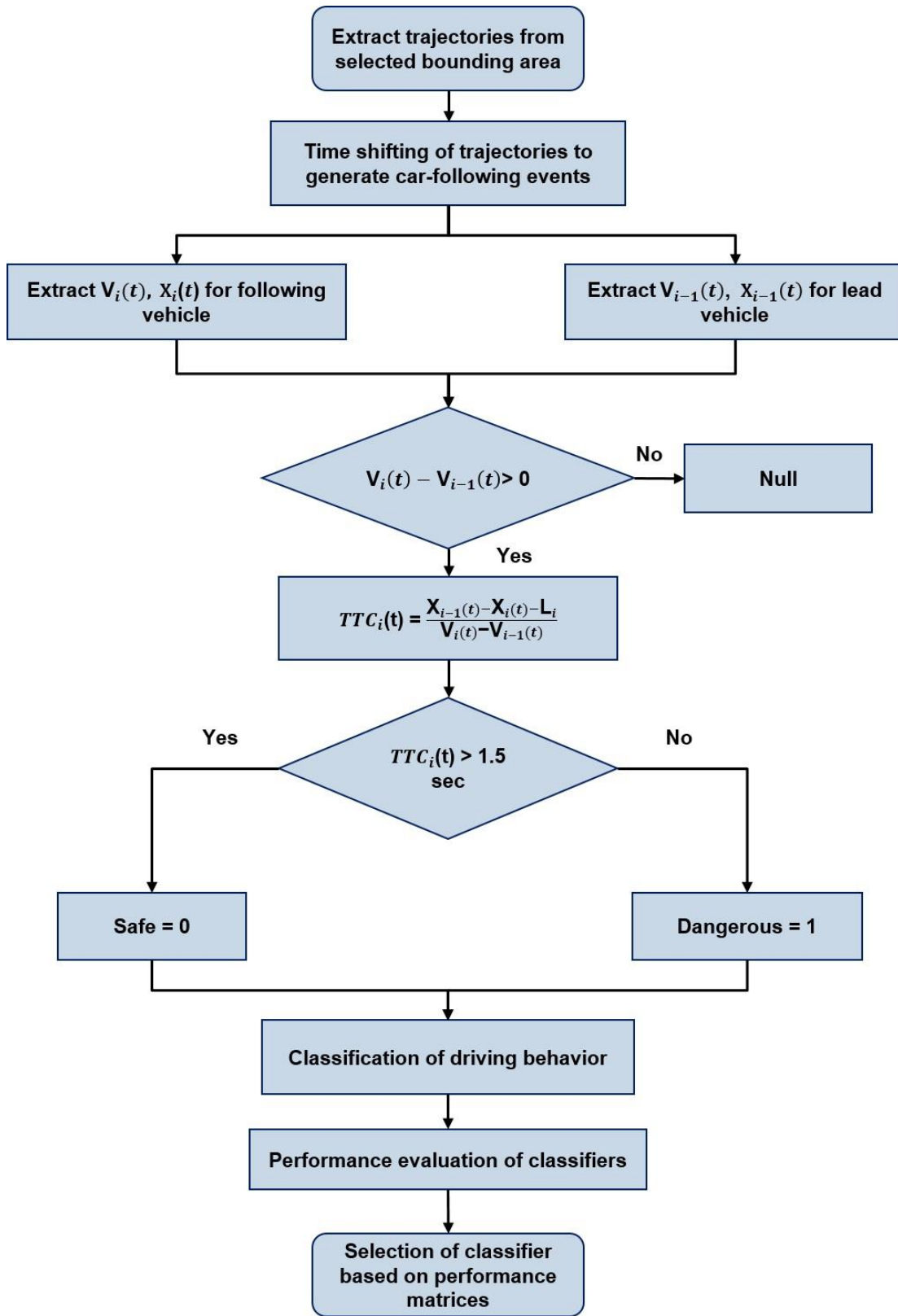
Figure 4.5 Chronological steps in Methodology

5 Results and Discussion

5.1 Identification of dangerous driving behaviour from TTC analysis

5.1.1 Change of TTC with distance to intersection

The T- Intersection is unsignalized and not controlled. From TTC calculation, it is observed that TTC value decreases when the vehicle approaches to the intersection. Figure 5.1 shows two pairs of trajectories in south-east and south-west manoeuvres, where TTC decreases with vehicle approaching to the intersection. Increase of TTC with distance from intersection can be correlated with an exponential growth curve.
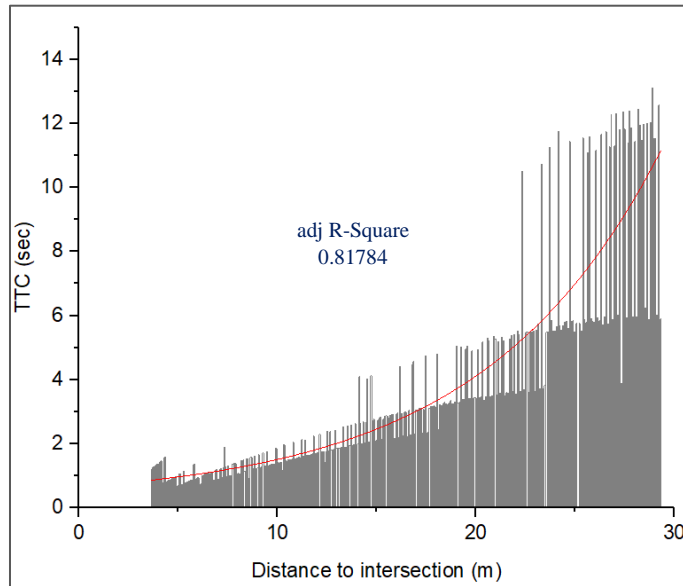
(a)
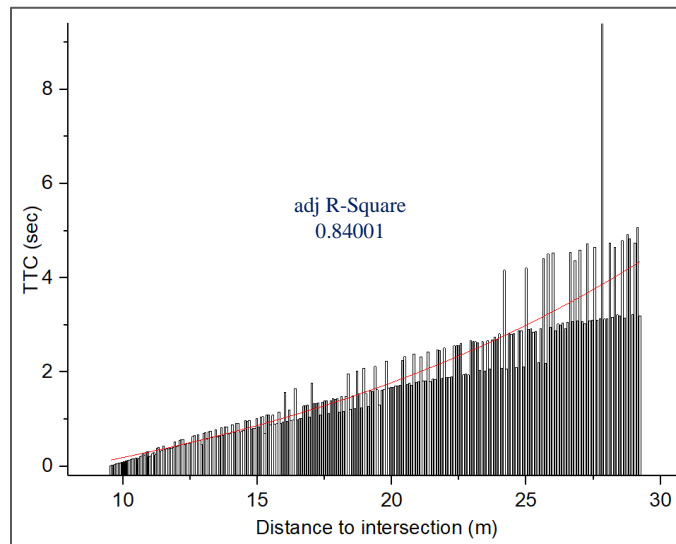
(b)

Figure 5.1 Decreasing trend of TTC approaching to intersection (a. south-east; b. South-west)

Pairs in other manoeuvres also show the same trend.

5.1.2 Change of TTC with speed difference

5.1.2.1 South-east manoeuvre

In south-east manoeuvre there were 10 trajectories. The first trajectory entering the bounding area is considered as the lead vehicle trajectory and was controlled and time shifted for calculating TTC with other trajectories. Figure 5.2 visualizes TTC from nine pairs of trajectories in south-east manoeuvre in boxplot with the mean value. It is obvious from the mean values of TTC that the first, fourth, fifth, sixth and seventh pairs are mostly collision-prone, where mean TTC value is less than 1.5 seconds.
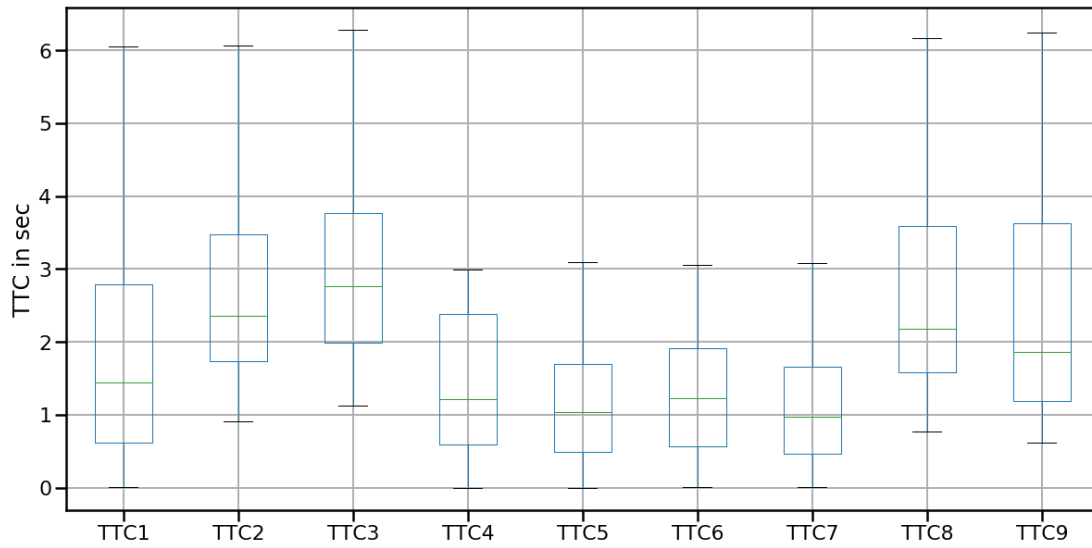
Figure 5.2 TTC in south-east manoeuvre

In this thesis, speed difference is taken as the independent variable in classifying dangerous driving behaviour. So, statistical analysis is done to see how TTC changes with speed difference between following vehicle and lead vehicle. High speed of the following vehicle leads to lower TTC or dangerous situation. The decrease in TTC is exponential with the increase in speed difference. Figure 5.3 shows an exponential decrease in TTC with increasing speed difference between following and lead vehicle in the first pair of trajectories at south-east direction.
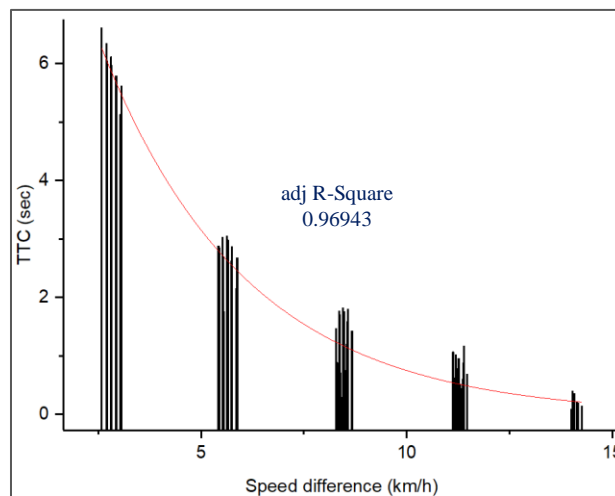
Figure 5.3 Change of TTC with speed difference for TTC1 in south-east manoeuvre

If y_t is TTC at time t, y_0 is initial TTC and λ is exponential decay constant; then decreasing trend of TTC can be expressed by the following first order exponential decay equation:

$$y_t = y_0 * \exp^{-\lambda t} \tag{5.1}$$

It is observed that for the scenarios, where speed difference is over 11 km/h, TTC comes below the threshold value of 1.5 seconds. To show the effect of speed difference on collision-prone situations, value of TTC is categorized in safe (=0) or dangerous (=1) as described in section 4.4. Figure 5.4 shows the mean value for speed difference in case of safe driving is less than 7 km/h and in dangerous driving it is almost 12 km/h. Cross mark in the boxplot shows the mean value.
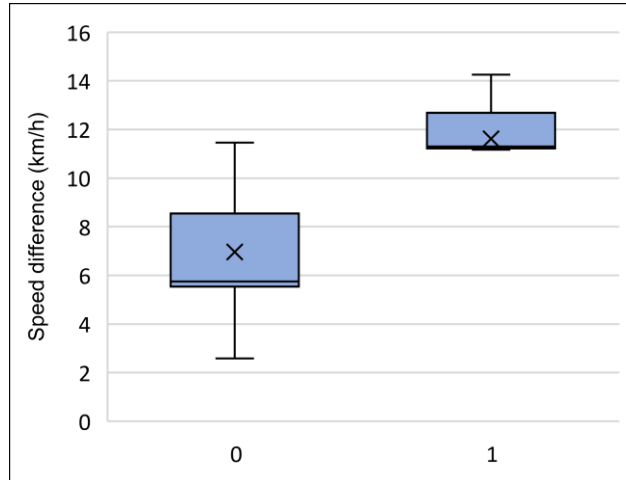
Figure 5.4 Speed difference for safe and dangerous situations for TTC1 in south-east manoeuvre

All other pairs of trajectories in south-east manoeuvre show the same trend of exponential decrease of TTC with increase of speed difference. After categorizing speed difference for all the pairs in Figure 5.5, it is observed that if speed difference between the two vehicles is kept below 7 km/h, there is less risk for collision.
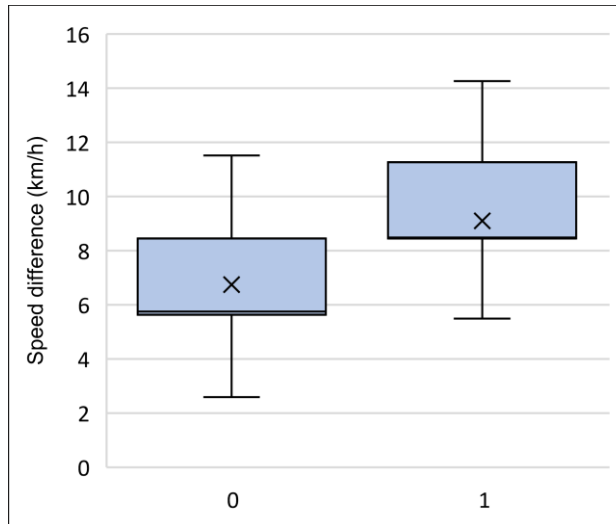
Figure 5.5 Speed difference for safe and dangerous situations for all pairs in south-east manoeuvre

5.1.2.2 East-south manoeuvre

The driver passed through the intersection nine times in east-south manoeuvre. Among them seven pairs of trajectories are extracted, because speed of the eighth trajectory has lower value than of the first one (e.g. lead vehicle speed is higher) and there is no chance of collision (as described in 4.3). Hence, there is no TTC7. Speed difference between lead and following vehicle in all the pairs are less in east-south manoeuvre. As a result, mean value of TTC is also lower and rate of dangerous situation is also very low (14.45%). Figure 5.6 visualizes TTC in the pairs of trajectories in east-south manoeuvre.
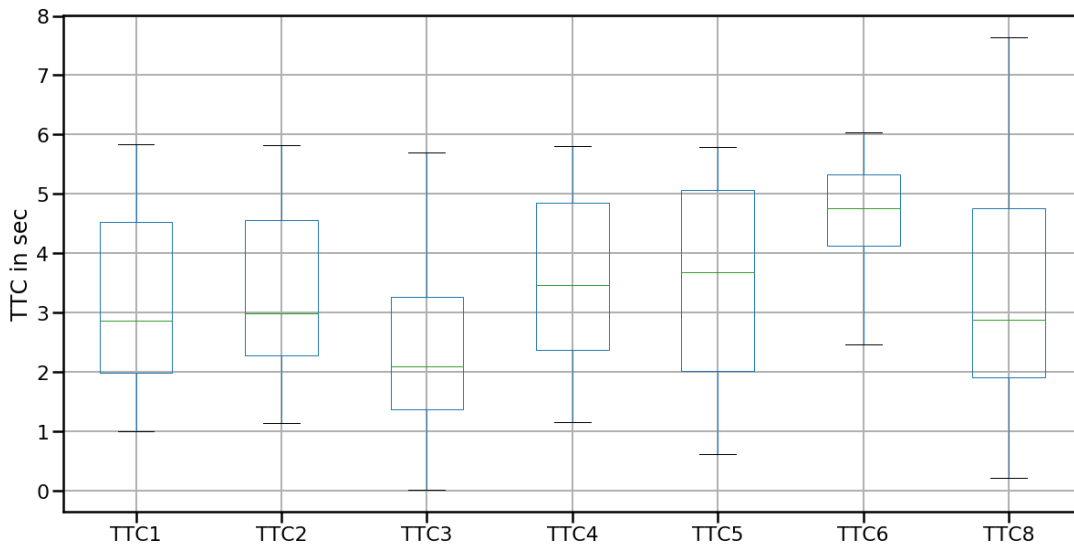
Figure 5.6 TTC in east-south manoeuvre

In east-south manoeuvre, change of TTC with speed difference follows the same trend as south-east manoeuvre. With increase in speed difference TTC decreases exponentially and collision risk increases (Figure 5.7 (a)). Mean value of speed difference is low for both safe and dangerous situations as shown in Figure 5.7 (b).

(a) (b)

Figure 5.7 Relationship between speed difference and TTC for safe and dangerous situations for all pairs in east-south manoeuvre

34

5.1.2.3 South-west manoeuvre

Nine pairs of trajectories were found from ten trajectories in south-west manoeuvres as shown in Figure 5.8. It is observed that in south-west manoeuvre, mean value of TTC in all the pairs are close to the threshold 1.5 seconds. That is why, half of the observations in south-west manoeuvre are found dangerous in TTC analysis (shown in Table 4.1).
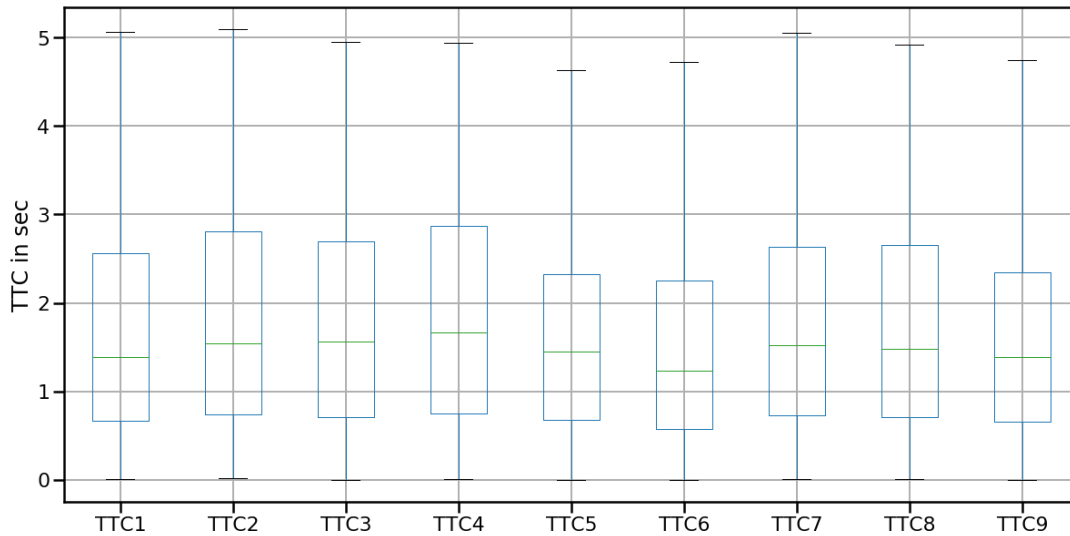
Figure 5.8 TTC in south-west manoeuvre

Figure 5.9 depicts that speed of the following vehicles are higher than the lead vehicle in south-west manoeuvre. As a result, mean value of speed difference for both safe and dangerous behaviour are also high (Figure 5.9(b)). Mean value of speed difference for dangerous behaviour is over 12.75 km/h in this manoeuvre.


(a)

(b)

Figure 5.9 Relationship between speed difference and TTC for safe and dangerous situations for all pairs in south-west manoeuvre

5.1.2.4 West-south manoeuvre

There are also nine pairs of trajectories found in west-south manoeuvre from car-following events. Rate of dangerous driving situation (TTC less than 1.5 seconds) found in west-south is 30.09%. Figure 5.10 shows the values of TTC in the trajectory pairs with mean value.
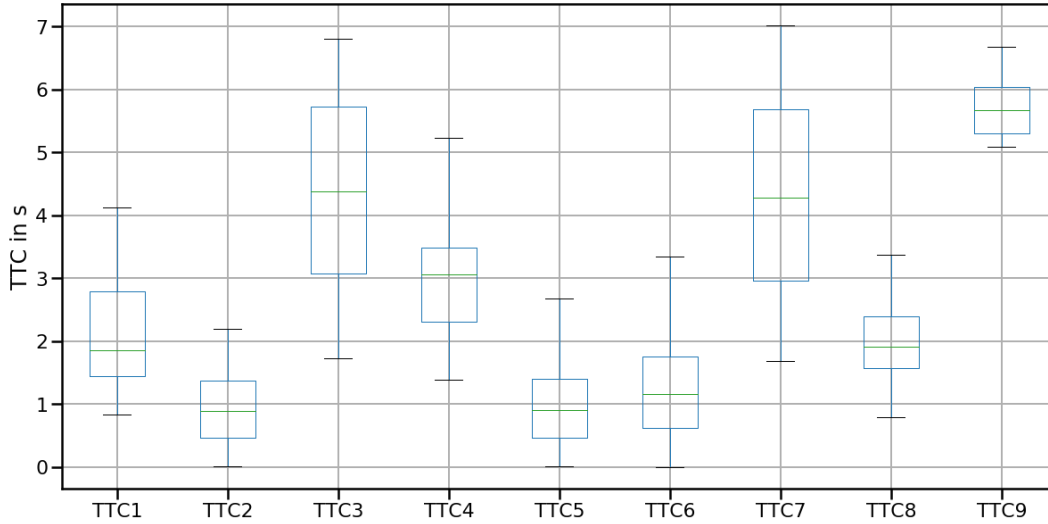
Figure 5.10 TTC in west-south manoeuvre

Mean value of speed difference for dangerous driving is over 7.2 km/h in west-south manoeuvre as shown in Figure 5.11 (b). For safe scenarios, the mean speed difference is 4.6 km/h.
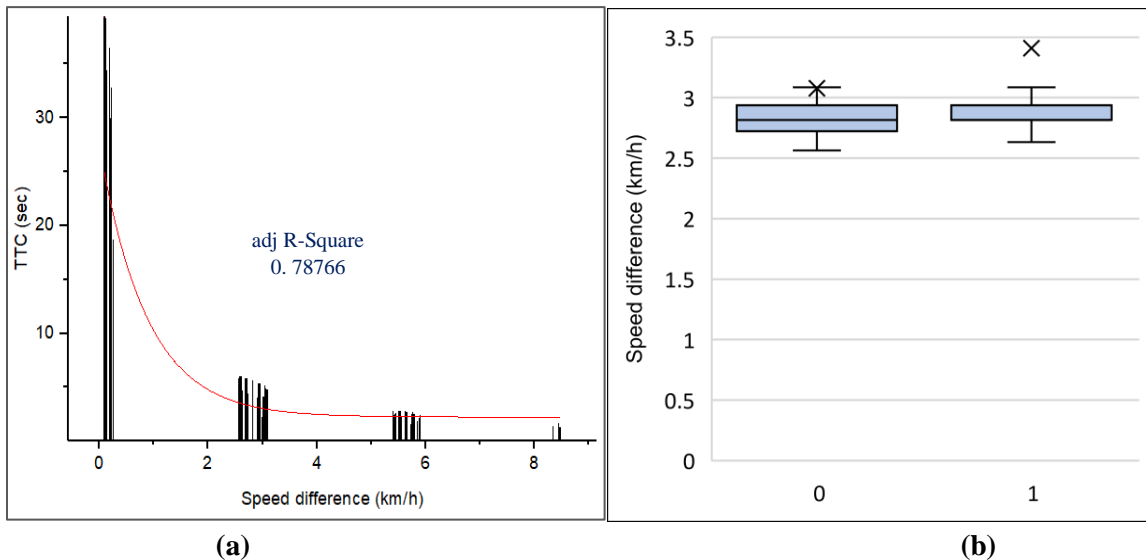

(a)   (b)
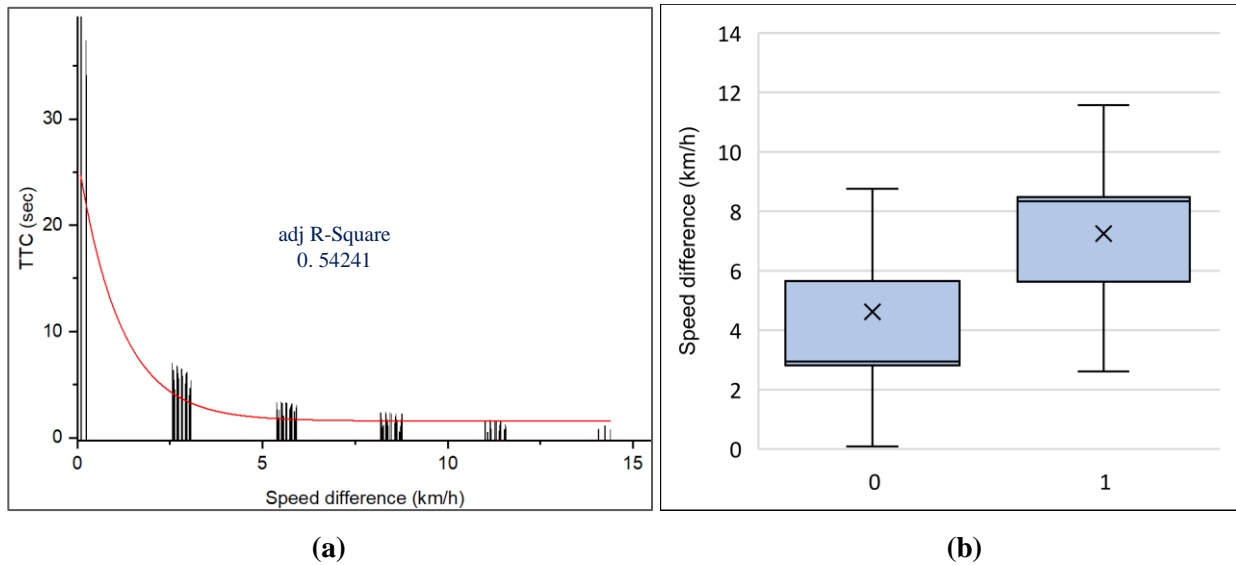
Figure 5.11 Relationship between speed difference and TTC for safe and dangerous situations for all pairs in west-south manoeuvre

5.1.2.5 East-west manoeuvre

From car-following events, eight pairs of trajectories are found in east-west manoeuvre because the ninth trajectory has less speed than the first vehicle trajectory and hence omitted in the car-following events. Rate of dangerous events, where TTC is below the threshold is 45.48% in this direction. TTC pairs in east-west manoeuvre are illustrated in Figure 5.12.
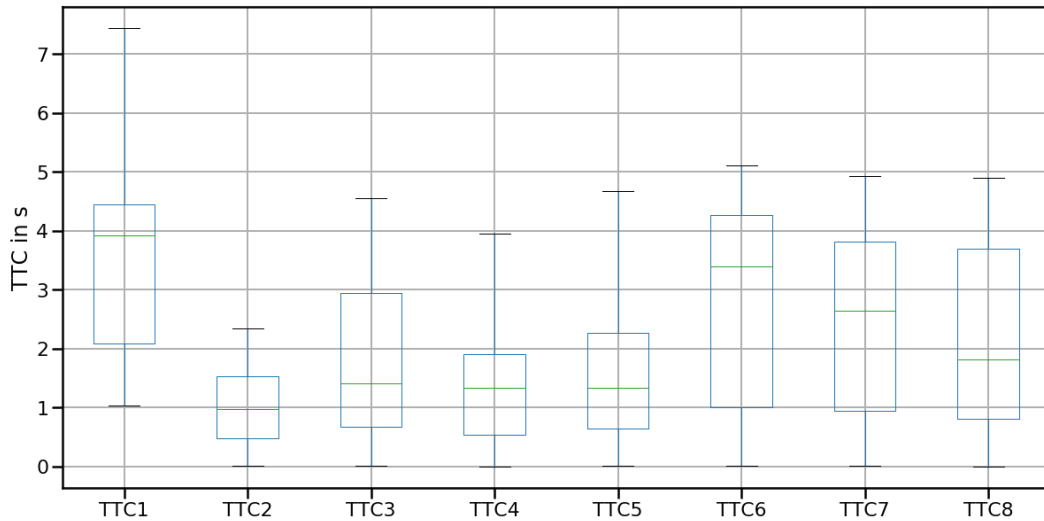
Figure 5.12 TTC in east-west manoeuvre

Figure 5.13 (a) shows exponential decrease in TTC with increase in speed difference between following and lead vehicle. From Figure 5.13 (b), it is observed that Mean value of speed difference for dangerous driving is 6.7 km/h while for safe driving the mean value is 3.9 km/h.

(a)   (b)

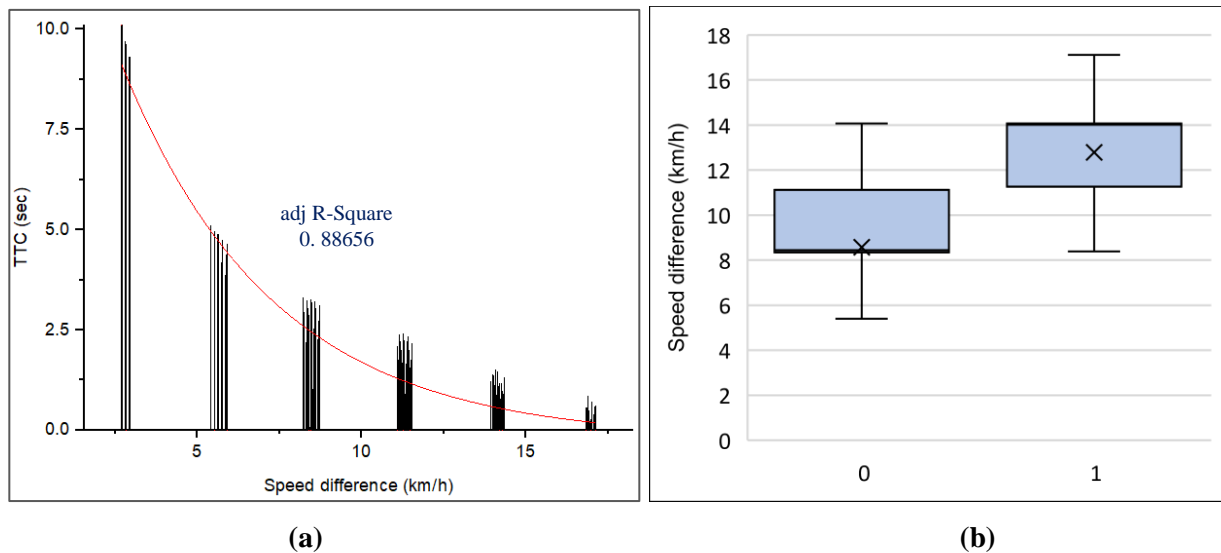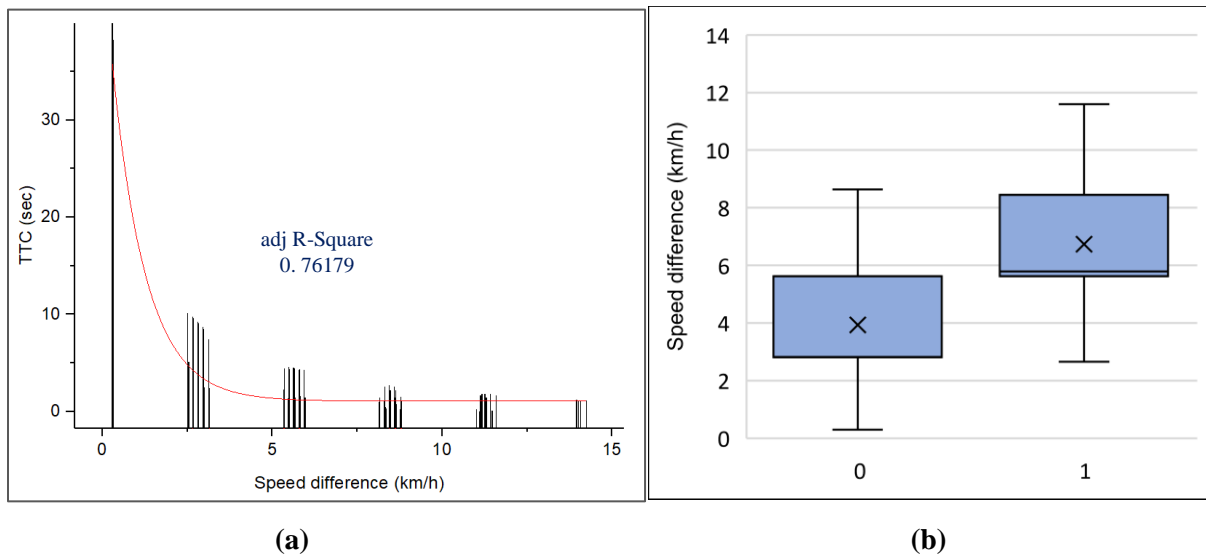**Figure 5.13** Relationship between speed difference and TTC for safe and dangerous situations for all pairs in east-west manoeuvre

5.1.2.6 West-east manoeuvre

In west-east manoeuvre, there is no dangerous driving situation found in the car-following events because in most of the cases speed of the following trajectories are less than the speed of first trajectory ($V_{i-1} > V_i$). So, TTC becomes negative and there is no collision risk. The mean speed difference in 247 car-following observations in this manoeuvre is only 2.88 km/hour (Figure 5.14).
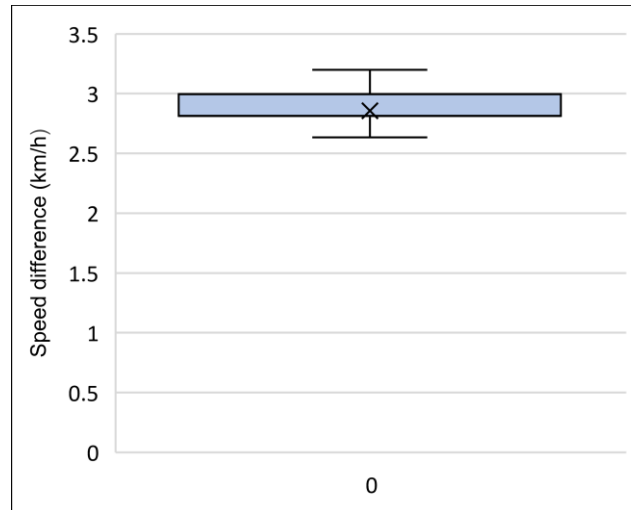
Figure 5.14 Speed difference in west-east manoeuvre

5.1.2.7 Visualization of speed difference for all observations

Figure 5.15 visualizes speed difference between following and lead vehicle for all 11,208 observations in six manoeuvres in the bounding area. Speed difference in safe situations is labelled as 0 and in dangerous situations as 1. It is obvious from the figure that high speed difference leads to dangerous situations. Mean value of speed difference for dangerous situations is 9 km/h, while for the safe situations mean value is less than 5.3 km/hour.
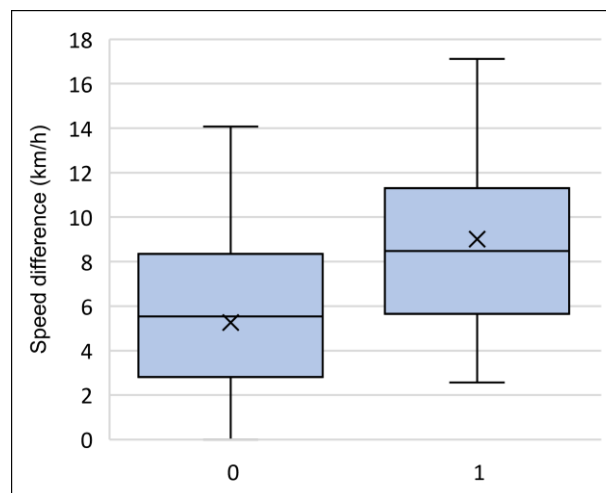
Figure 5.15 Speed difference for safe and dangerous situations for all observations in the bounding area

5.1.3   Clustering of TTC and speed difference

Clustering is an unsupervised machine learning technique, which is opposite to classification and considered as a grouping technique. Clustering has been used in identifying collision risk for vehicles by different researchers. Zheng et al. (2014) used three clusters of deceleration values to identify high, moderate and low risk of collision from a naturalistic driving dataset. K-means clustering is a very popular technique, which is mainly used to find association between data based on inherent similarities between them (Castle 2017). Number of unique groups is represented by K. Association among data is identified by the shortest distance of data from centroid of clusters. Guo & Fang (2013) used k-means clustering technique to assess individual driving risk using naturalistic data.

The aim of this thesis is to identify dangerous driving behaviour. So, k-means clustering can be used to visualize the grouping between safe and dangerous driving behaviour. Based on TTC analysis in all the manoeuvres from section 5.1.2, two groups or clusters can be employed to understand the association between speed difference and TTC. Figure 5.16 represents the association among data in all the manoeuvres. Values in the yellow cluster represents high TTC or safe situation and blue cluster values represent low TTC e.g. dangerous situation. Centroid of the clusters are represented by red dots. It is evident from the graph that high speed difference values lie in the blue cluster which define dangerous driving behaviour and low speed difference values lie with the values in yellow cluster or safe behaviour. K-means clustering is a good technique to represent anomaly in the data which affect the performance of classification. Figure 5.16 also shows some anomaly when low speed difference represents low TTC and vice-versa.
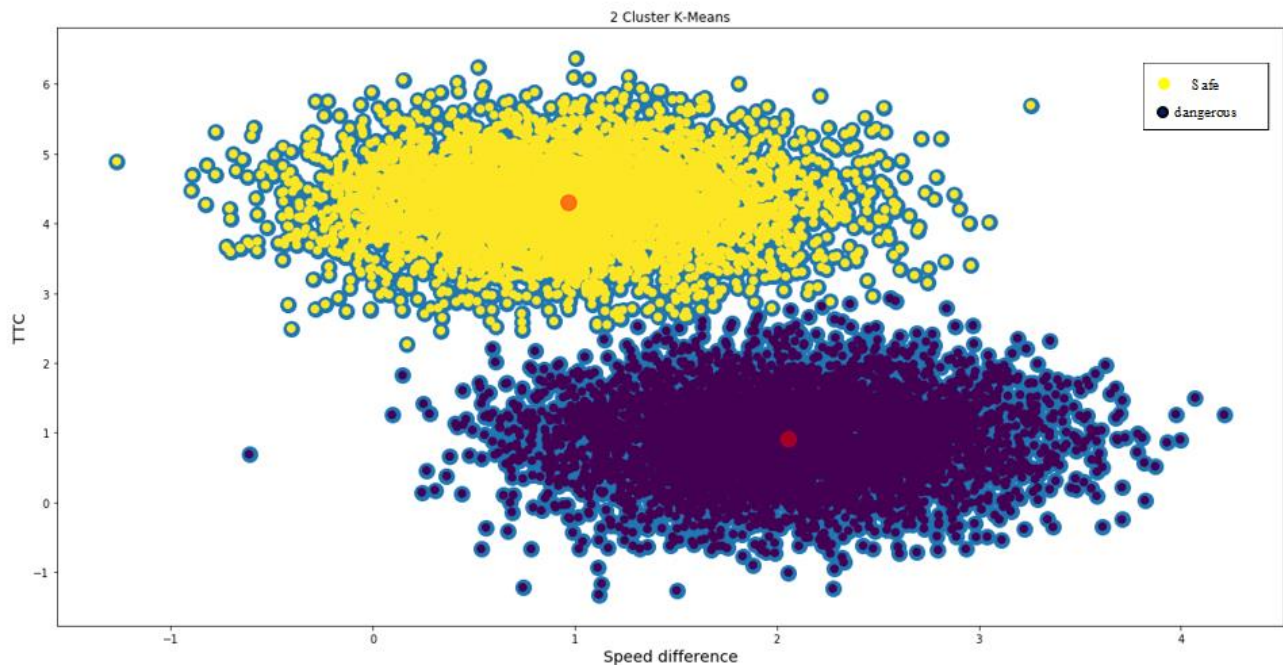
Figure 5.16 K-means clustering between TTC and speed difference

5.2 Comparison of classifiers

As described in section 4.4, seven machine learning classifiers (Gaussian Naive Bayes, Bernoulli Naive Bayes, Logistic Regression, Support Vector Machine, Decision Tree, k-Nearest Neighbor and Random Forest) are compared to classify or predict dangerous driving behaviour in Python 3.7 by using sci-kit learn package (Pedregosa et al. 2011). Speed difference is taken as the predictor and binary outcome 0 or 1 stands for safe or dangerous situations respectively.

Overfitting is a common problem in machine learning classification and prediction performance is reduced due to this problem. An overfitted model contains more parameters that can be justified by the dataset (Everitt and Skrondal 2010). To avoid this problem, cross-validation is widely used to evaluate machine learning models. K-fold cross validation is a popular cross-validation technique to estimate performance of classifiers on unseen data. This technique splits the dataset into k number of groups for less bias in classification (James et al. 2013).

In this thesis, 10-fold cross-validation was run on the dataset before employing each classifier in order to find the optimal parameters for each algorithm. Moreover, to avoid overfitting, 2/3 of the datasets were used for training and the remaining l/3 was used for testing. The classifiers are evaluated by some performance measures derived from the confusion matrix (discussed in 4.5.1). Confusion matrix for seven different classifiers is given in Table 5.1.

Classifier	TN	FP	FN	TP	
Gaussian Naive Bayes	2184	169	781	565	
Bernoulli Naive Bayes	2353	0	1346	0	Testing sample size=3699 (safe=2353, dangerous= 1346)
Logistic Regression	2169	184	773	573	
Support Vector Machine	2059	294	698	648	
k-Nearest Neighbor	2094	259	552	794	
Decision Tree	2191	162	632	714	
Random Forest	2190	163	628	718	

Table 5.1 Confusion matrix of the classifiers

It is observed from Table 5.1 that False positive or false danger prediction is the least in number for Decision Tree and Random Forest classifier compared to other classifiers. Different performance metrices obtained from confusion matrix are used to evaluate the performance of the classification techniques. Table 5.2 shows the result of different performance matrices derived from the confusion matrix Table 5.1.

Classifier name	Accuracy	Precision (%)	Recall (%)	Specificity (%)	f1-score (%)	False alarm rate (%)
Gaussian Naive-Bayes	74.317	76.975	41.976	92.817	54.327	7.18
Bernoulli Naive-Bayes	63.612	NA	0	100	-	-
Logistic Regression	74.128	75.693	42.57	92.18	54.49	7.82
Support Vector Machine	73.182	68.789	48.14	87.505	56.64	12.49
k-Nearest Neighbor	78.075	75.403	58.98	88.99	66.19	11.007
Decision Tree	78.534	81.50	53.04	93.11	62.266	6.88
Random Forest	78.669	81.50	53.34	93.073	64.481	6.92

Table 5.2 Performance evaluation of different classifiers

Table 5.2 depicts that all the classifiers show high percentage of specificity compared to recall and f1-score. That demonstrates that the classifiers can distinguish safe driving behaviour easily but cannot detect dangerous behaviour in the same rate. It is obvious from the performane matrices that Decision Tree and Random Forest classifier give the best prediction of dangerous driving situations among the seven classifiers tested, based on better accuracy, f1-score and least rate of false collision prediction.

Figure 5.17 visualizes the comparison of different classifiers based on accuracy, f1-score and false alarm rate.
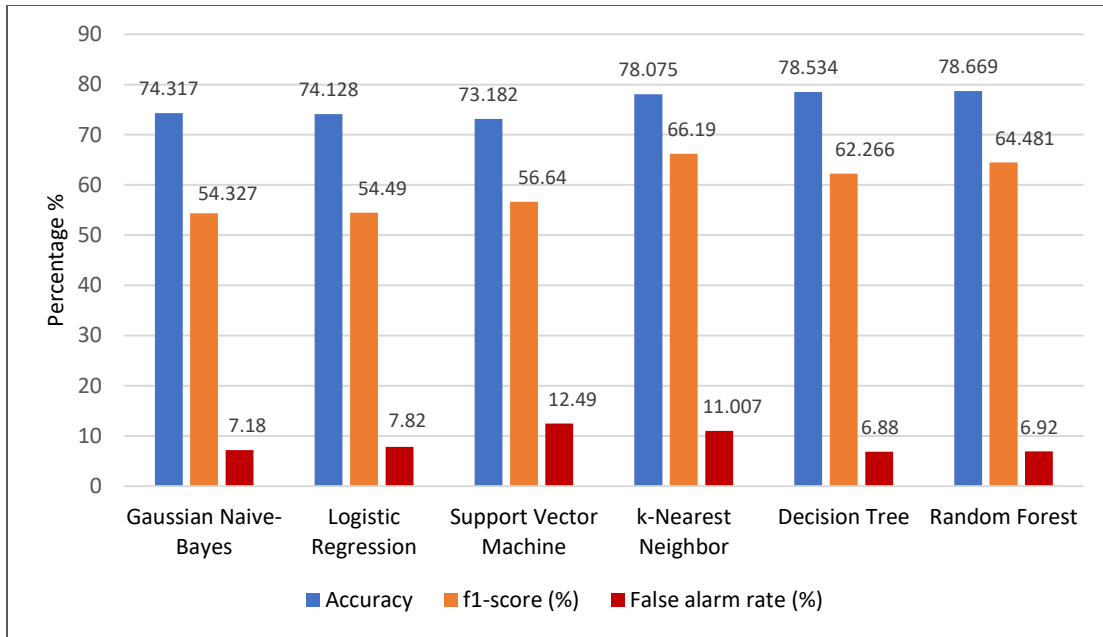
Figure 5.17 Comparison of classifiers based on accuracy, f1-score and false alarm rate

Area under ROC curve (auc) is a good measure to visualize the tradeoff between false positive rate and true positive rate. Figure 5.18 visualizes auc value under ROC curve for different classifiers.
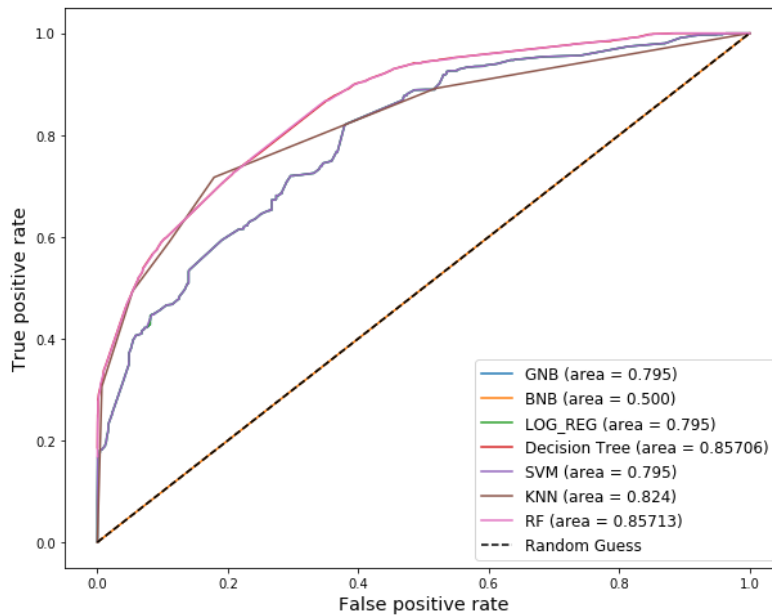
Figure 5.18 ROC curve of different classifiers

5.3   Classifying driving behaviour by Random Forest classifier

In this thesis, Random Forest classifier is used to predict dangerous driving behaviour because it gives better area under ROC curve and also better f1-score which is an increasing function for both precision and recall. When employing RF in all the manouevres, 2/3 of the dataset is used for training and 1/3 of the dataset is used as test samples in order to avoid overfitting. Table 5.3 shows the confusion matrix of Random Forest classifiers in all manoeuvres.

Manoeuvre	TN	FP	FN	TP	Testing sample size
South-east	511	61	169	190	931(safe=572, dangerous=359)
East-south	539	1	83	10	633(safe=540, dangerous=93)
South-west	327	47	16	402	792(safe=374, dangerous=418)
West-south	358	43	67	97	565(safe=401, dangerous=164)
East-west	285	104	37	272	698(safe=389, dangerous=309)
West-east	-	-	-	-	-

Table 5.3 Confusion matrix of RF in all manoeuvres

Table 5.4 shows evaluation results of Random Forest classifier by performance matrices for predicting dangerous driving behaviour in all the manoeuvres for driver 3.

Manoeuvres	Accuracy (%)	Precision (%)	Recall (%)	Specificity (%)	f1-score (%)	False alarm rate (%)
South-east	75.29	76.11	52.37	89.69	62.05	10.31
East-south	86.73	90.91	10.75	99.81	19.23	0.18
South-west	92.04	89.53	96.17	87.43	92.73	12.57
West-south	80.71	68.97	60.98	88.78	64.72	11.22
East-west	79.94	72.41	88.35	73.26	79.26	26.74
West-east	-	-	-	-	-	-

Table 5.4 Performance matrices of Random Forest classifier in all manoeuvres

Table 5.4 shows that Random Forest classifier performs well in predicting dangerous driving behaviour for majority of the manoeuvres except east-west where false alarm rate is 26.74%. Percentage of recall is very low in east-south manoeuvre because of the high amount of imbalanced data in test set where number of dangerous situations is very few (only 93) compared to 540 safe situations (from Table 5.3). High number of specificity in all the manoeuvres indicate that the model performs better in identifying safe driving behaviour than the dangerous behaviour. High percentage of precision in the manoeuvres (except west-south) demonstrates that when the model identifies a dangerous driving situation, it is more likely to be dangerous scenario rather than false alarm or safe situation. Figure 5.19 shows performance of Random Forest classifier in different manoeuvres based on accuracy, f1-score and false alarm rate.
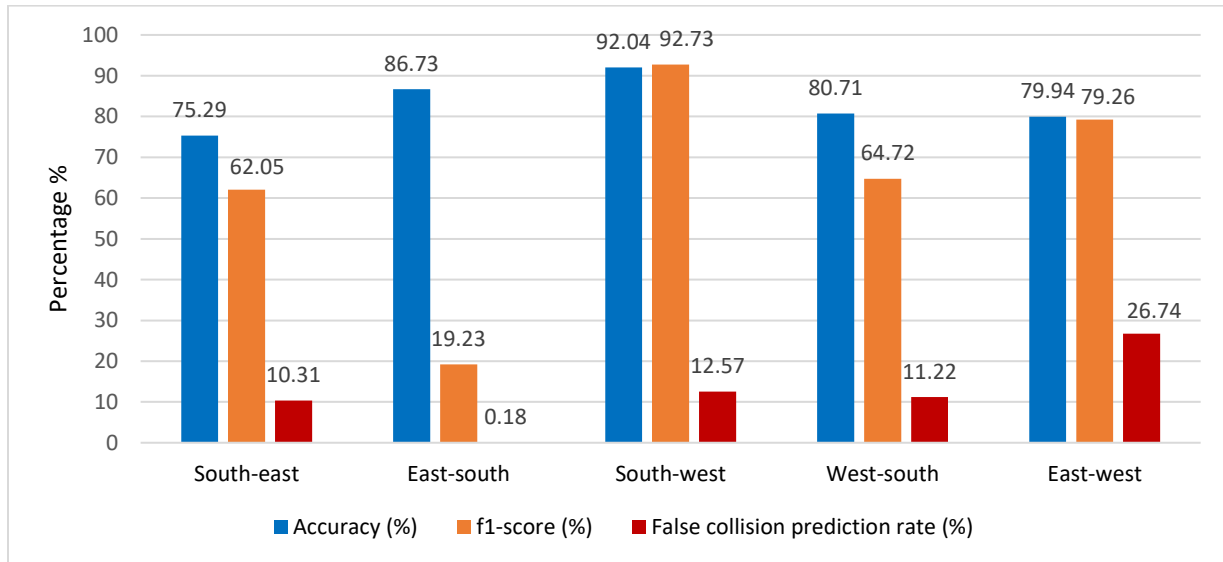
Figure 5.19 Accuracy, f1-score and false alarm rate of Random Forest for different manoeuvres

Area under ROC curve for RF in all the manoeuvres is shown in Figure 5.20. Area under ROC curves indicate that RF classifier performs well in predicting the dangerous situations in different manoeuvres apart from east-south because of high imbalance among data between safe and dangerous driving behaviour.
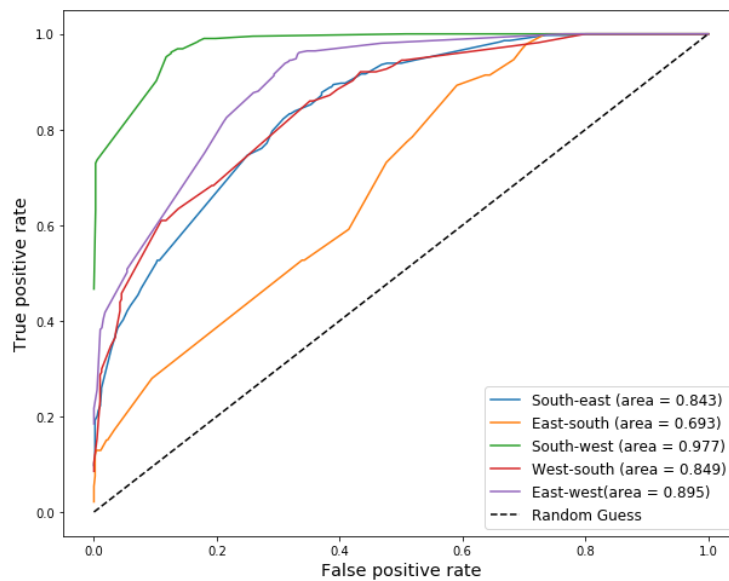
Figure 5.20 ROC curve of Random Forest classifier for different manoeuvres

5.4 Results summary and discussion

This thesis aims at identifying and classifying dangerous driving behaviour for autonomous vehicles from a naturalistic driving dataset. The dataset contains only single vehicle's position, speed and heading data. Trajectories are extracted from position data and time shifted to generate collision or dangerous situations around the uncontrolled T-intersection. The bounding area was selected as 35 meters from the center of T-intersection in every direction.

Surrogate safety measure TTC is applied to identify safe or collision prone situation. In total, 35.96% dangerous situations are found from 11,208 car-following observations in all manoeuvres, where TTC lies below the threshold of 1.5 seconds. It is observed that the decrease in TTC with the increasing speed difference is exponential. In general, if speed difference between following vehicle and lead vehicle is more than 2.5 m/s or 9 km/h, dangerous situation arises. TTC value comes lower when the vehicles approach to the intersection. K-means clustering is used to show the association between TTC and speed difference by visualizing two groups as safe and dangerous driving behaviour.

Finally, seven machine learning classification algorithms are tested on all the manoeuvres to classify dangerous driving behaviour. Speed difference is considered as the only predictor for classifying dangerous situations. In order to avoid overfitting and reduce bias in classification, 10-fold cross-validation was run on the data before employing the classifiers and l/3 of the data samples were used for testing. Different performance matrices are used to evaluate the classifiers and Random Forest classifier comes out to perform better in predicting dangerous situations compared to other classifiers. It gives only 6.92% false collision alarm and outperforms other classifiers in terms of f1-score. Area under the curve of ROC also increases for RF classifier. However, in some manoeuvres, it does not identify dangerous behaviours in a better rate because of either high imbalance of data between safe and dangerous situations or small test sample.

According to Lefèvre (2012), evaluation of vehicle level risk assessment models should be done on the same dataset. Hence, the performance of the classification technique used in this thesis is not compared with existing approaches from literature. Random Forest is primarily used for variable selection in existing literature (Hassan and Abdel-Aty 2013, Hossain and Muromachi 2013, Xu et al. 2013, Ahmed and Abdel-Aty 2012). However, in this thesis it gives more robust performance than other widely used classifiers in classifying driving behaviour.

Literature suggests that imbalanced learning techniques enhance the collision prediction performance of classifiers, when imbalance exists between 'safe' and 'collision-prone' samples (Katrakazas 2017, Katrakazas et al. 2019). It is believed that integration of imbalanced learning techniques like oversampling of dangerous situation samples, under sampling of safe samples or ensemble learning techniques with Random Forest classifier would have provided more accuracy in predicting dangerous driving behaviour.

6 Conclusion

6.1 Summary

Traffic accidents are still one of the dominant reasons for huge amount of human deaths/injuries and economic loss in present world. Mostly, misperception and decision making from human drivers are responsible for collision occurrences. Introduction of autonomous vehicles has been proven to be a big advancement in road safety research for mitigating the error in human driving. Research is still going on for finding accurate risk assessment models of autonomous vehicles.

Motion planning is the prerequisite for safe navigation of autonomous vehicles which defines the safest manoeuvres for driving through the trajectories according to traffic dynamics. However, existing literature of motion planning and risk assessment approaches lack proper AV applications and face the complexity of surrounding traffic participants and environment. Most of the approaches overlook interaction among traffic participants and consider them as independent entities. In recent times, improvement in ITS and advanced data collection technologies have facilitated traffic safety research immensely. Real-time collision prediction is a big step in road safety analysis which classifies present traffic condition into safe or collision-prone by comparing traffic conditions at normal situation and just before the collision occurs. Existing literature mostly describes the real-time collision prediction based on road network level data which is defined as Network Level Collision Prediction (NLCP). However, vehicle level collision prediction is also required for understanding driving behaviour and vehicle dynamics in safety assessment of autonomous vehicles.

This thesis aimed at identifying dangerous driving behaviour of AVs for rear-end collision in an uncontrolled three-way T-intersection by using a naturalistic driving data. Only single vehicle data is used in the experiment to understand vehicle level collision prediction and to interpret how the speed changes at times of collision. Trajectories were time shifted and controlled to have interactions among them for rear-end collision scenario. Surrogate safety indicator TTC was used to identify dangerous situation and how speed variation affect safe or collision-prone situations in different manoeuvres.

Furthermore, different machine learning classification techniques were applied to classify safe or dangerous speed difference in rear-end collision scenario. Seven classifiers were tested, and Random Forest classifier performed the best in predicting dangerous behaviour after evaluation. Combined false alarm rate is less than 7% for RF in predicting dangerous driving situations. However, in some manoeuvres it does not predict well because of high imbalance between safe and collision-prone samples. It is expected that, more sophisticated real-world traffic data and integration of more advanced classification techniques like imbalanced learning, deep learning with this model will provide better classification performance.

6.2　Limitations and future work

This research aimed at classifying dangerous driving behaviour by identifying them from time-shifted trajectories from a naturalistic driving dataset, which contains single vehicle data from sensor measurements. Machine learning algorithms were employed for predicting safe or dangerous behaviour by selecting only one variable speed difference. Classification results could be improved if more predictors like vehicle acceleration/deceleration or yaw rate were available in dataset.

As trajectories are time-shifted and car-following events are controlled artificially for rear-end collision, there is a little imbalance in the number of total safe and collision-prone situations. However, in real-world, the data samples are highly imbalanced and there is high number of safe samples than collision-prone samples which can derail the classification performance of classifiers.

In some manoeuvres, false alarm rate is higher, because of the imbalance among data samples. In that case, it is believed that imbalanced learning like oversampling of collision-prone samples and under sampling of safe samples would have given better prediction. It is suggested to integrate the imbalanced learning techniques with Random Forest classifiers for better accuracy in classification.

Only rear-end collision scenarios in the intersection is studied in this thesis. Classification of dangerous driving behaviour for turning and head-on collisions by time-shifting of trajectories will be a scope of research in future.

The naturalistic driving study for data collection was performed only on single-lane urban streets and the road feature is a three-way T-intersection. The classification technique applied in this thesis can be tested on double-lane expressways with four-way intersection, to include lane changing behaviour in rear-end collision risk assessment.

Although, position data were recorded at every 10 Hertz by GNSS, there was measurement noise in data. The noise in the data measurement is shown in one of the collected trajectories in Figure 6.1. A more robust risk assessment model can be achieved by processing the data for recovering smooth trajectories (Ward et al. 2014b, Agamennoni et al. 2010).
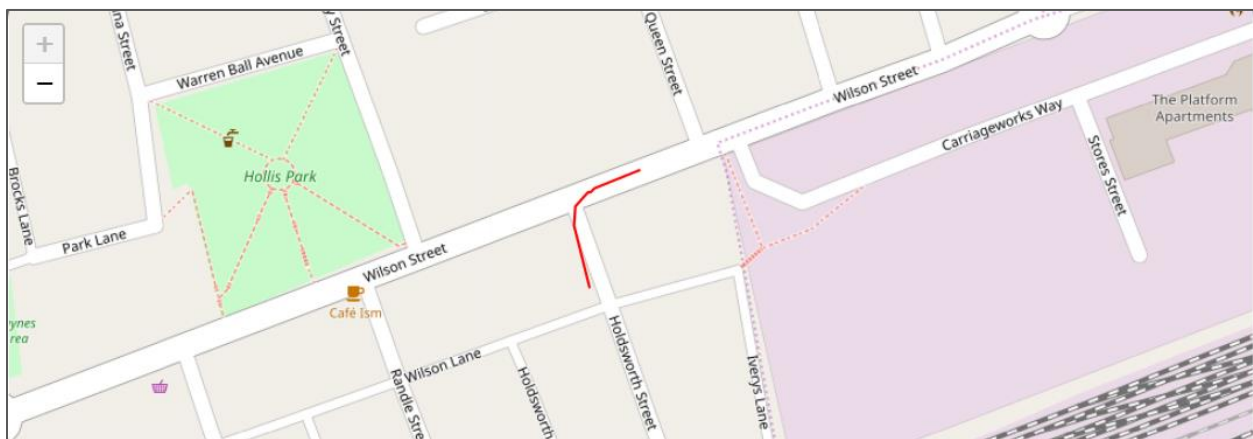

Figure 6.1 Noise in measurements in one of the trajectories

Finally, only vehicle level collision prediction is considered in this research. The integration of network level collision prediction with this classification method will be valuable for the safety assessment of autonomous vehicles in future.

References

Abbas MH, Khan M-u-R (2007) Correlational analysis of driver's personality traits and styles in a distributed simulated driving environment. Institutionen för datavetenskap. p. 128.

Abdel-Aty M, Pande A (2005) Identifying crash propensity using specific traffic speed conditions. Journal of safety Research 36(1):97–108.

Abdel-Aty M, Uddin N, Pande A (2005) Split models for predicting multivehicle crashes during high-speed and low-speed operating conditions on freeways. Transportation Research Record (1908):51–58.

Abdel-Aty M, Uddin N, Pande A, Abdalla MF, Hsia L (2004) Predicting Freeway Crashes from Loop Detector Data by Matched Case-Control Logistic Regression. Transportation Research Record 1897(1):88–95.

Agamennoni G, Nieto JI, Nebot EM (2010) Robust and accurate road map inference. In: 2010 IEEE International Conference on Robotics and Automation. Anchorage, Alaska, May 3-8, 2010: 50 years of robotics. IEEE, [Piscataway, N.J.], pp 3946–3953.

Agamennoni G, Nieto JI, Nebot EM (2012) Estimation of Multivehicle Dynamics by Considering Contextual Information. IEEE Trans. Robot. 28(4):855–870.

Ahmed MM, Abdel-Aty M, Yu R (2012) Bayesian Updating Approach for Real-Time Safety Evaluation with Automatic Vehicle Identification Data. Transportation Research Record 2280(1):60–67.

Ahmed MM, Abdel-Aty MA (2012) The Viability of Using Automatic Vehicle Identification Data for Real-Time Crash Prediction. IEEE Trans. Intell. Transport. Syst. 13(2):459–468.

Alin A, Butz MV, Fritsch J (2012) Incorporating environmental knowledge into Bayesian filtering using attractor functions. In: IEEE Intelligent Vehicles Symposium (IV), 2012. 3-7 June 2012, Alcalá de Henares, Madrid, Spain. IEEE, Piscataway, NJ, pp 476–481.

Allen BL, Shin BT, Cooper PJ (1978) Analysis of traffic conflicts and collisions. Road and Motor Vehicle Safety Branch, Transport Canada, Ottawa.

Ammoun S, Nashashibi F (2009) Real time trajectory prediction for collision risk estimation between vehicles. In: Letia IA (ed) Proceedings. IEEE, Danvers, pp 417–422.

Aoude GS, Desaraju VR, Stephens LH, How JP (2012) Driver Behavior Classification at Intersections and Validation on Large Naturalistic Data Set. IEEE Trans. Intelligent Transportation Systems. 13(2):724–736.

Aoude GS, Luders BD, Levine DS, How JP (2010) Threat-aware path planning in uncertain urban environments. In: The IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems (IROS 2010). Taipei International Convention Center, Taipei, Taiwan, October 18-22, 2010: conference proceedings. Institute of Electrical and Electronics Engineers, Piscataway, N.J., pp 6058–6063.

Archer J (2005) Indicators for traffic safety assessment and prediction and their application in micro-simulation modelling. A study of urban and suburban intersections. Doctoral Thesis, KTH Royal Institute of Technology, Sweden.

Archer J, Young W (2010) Traffic Microsimulation Approach to Estimate Safety at Unsignalized Intersections. In: Transportation Research Board 89th Annual Meeting. Washington DC, USA.

References

Ariza A (2011) Validation of Road Safety Surrogate Measures as a Predictor of Crash Frequency Rates on a Large-Scale Microsimulation Network. Master Thesis. University of Toronto.

Armand A, Filliat D, Ibanez-Guzman J (2014) Ontology-Based Context Awareness for Driving Assistance Systems. In: IEEE Intelligent Vehicles Symposium. 2014. pp 227–233.

Asiri S (2018) Machine Learning Classifiers – Towards Data Science. Available online at https://towardsdatascience.com/machine-learning-classifiers-a5cc4e1b0623.

Astarita V, Giofré V, Guido G, Vitale A (2011) Investigating road safety issues through a microsimulation model. Procedia - Social and Behavioral Sciences 20:226–235.

Bahram M, Hubmann C, Lawitzky A, Aeberhard M, Wollherr D (2016) A Combined Model- and Learning-Based Framework for Interaction-Aware Maneuver Prediction. IEEE Trans. Intelligent Transportation Systems 17(6):1538–1550.

Balas VE, Balas MM (2006) Driver assisting by inverse time to collision. In: 2006 World Automation Congress, pp 1–6.

Bandyopadhyay T, Won KS, Frazzoli E, Hsu D, Lee WS, Rus D (2013) Intention-Aware Motion Planning. In: Frazzoli E (ed) Algorithmic foundations of robotics X. Proceedings of the Tenth Workshop on the Algorithmic Foundations of Robotics, vol 86. Springer, Heidelberg, New York, pp 475–491.

Bender A, Ward JR, Worrall S, Nebot EM (2015) Predicting Driver Intent from Models of Naturalistic Driving. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC). 15 - 18 Sept. 2015, Las Palmas, Gran Canaria, Spain. IEEE, Piscataway, NJ, pp 1609–1615.

Berwick R (2003) An Idiot's guide to Support vector machines (SVMs). Massachusetts Institute of Technology, USA.

Bishop CM (2006) Pattern recognition and machine learning. Springer New York.

Brechtel S, Gindele T, Dillmann R (2014) Probabilistic decision-making under uncertainty for autonomous driving using continuous POMDPs. In: Intelligent Vehicles Symposium Proceedings, 2014 IEEE, pp 392–399.

Broggi A, Medici P, Zani P, Coati A, Panciroli M (2012) Autonomous vehicles control in the VisLab Intercontinental Autonomous Challenge. Annual Reviews in Control 36(1):161–171.

Buehler M, Iagnemma K, Singh S (2009) The DARPA urban challenge: autonomous vehicles in city traffic, vol 56. Springer.

Castle N (2017) An Introduction to Machine Learning Algorithms. Available online at https://www.datascience.com/blog/introduction-to-machine-learning-algorithms.

Ceunynck T de (2017) Defining and applying surrogate safety measures and behavioural indicators through site-based observations. Hasselt University; Lund University, Diepenbeek, Lund.

Chin H-C, Quek S-T (1997) Measurement of traffic conflicts. Safety Science 26(3):169–185.

Dingus TA, Klauer SG, Neale VL, Petersen A, Lee SE, Sudweeks J, Perez MA, Hankey J, Ramsey D, Gupta S, others (2006) The 100-Car Naturalistic Driving Study. Phase 2: Results of the 100-Car Field Experiment.

Dolgov D (2016) Google self-driving car project monthly report. Published September.

References

Dolgov D, Thrun S, Montemerlo M, Diebel J (2010) Path Planning for Autonomous Vehicles in Unknown Semi-structured Environments. The International Journal of Robotics Research 29(5):485–501.

Dreiseitl S, Ohno-Machado L (2002) Logistic regression and artificial neural network classification models: a methodology review. Journal of Biomedical Informatics 35(5-6):352–359.

Eskandarian A (2012) Handbook of intelligent vehicles. Springer reference. Springer, London

Evans L (1991) Traffic safety and the driver. Van Nostrand Reinhold, New York.

Evans L (2004) Traffic safety. Science Serving Society of Bloomfield Hills, Michigan.

Evans L, Wasielewski P (1983) Risky driving related to driver and vehicle characteristics. Accident Analysis & Prevention 15(2):121–136.

Everitt B, Skrondal A (2010) Standardized mortality rate (SMR). The Cambridge Dictionary of Statistics 409.

Fagnant DJ, Kockelman K (2015) Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice 77:167–181.

Farber B (1991) Designing a distance warning system from the user point of view. APSIS report, Glonn-Haslach: Institute fur Arbeitspsychologie and Interdisziplinare Systemforchung.

Furda A, Vlacic L (2011) Enabling Safe Autonomous Driving in Real-World City Traffic Using Multiple Criteria Decision Making. IEEE Intell. Transport. Syst. Mag. 3(1):4–17.

Gadepally VN (2013) Estimation of driver behavior for autonomous vehicle applications. Doctoral Thesis. The Ohio State University.

Gahukar G (2018) Classification Algorithms in Machine Learning. Available online at https://medium.com/datadriveninvestor/classification-algorithms-in-machine-learning-85c0ab65ff4

Gettman D, Head L (2003) Surrogate Safety Measures from Traffic Simulation Models. Transportation Research Record 1840(1):104–115.

Gindele T, Brechtel S, Dillmann R (2015) Learning Driver Behavior Models from Traffic Observations for Decision Making and Planning. IEEE Intell. Transport. Syst. Mag. 7(1):69–79.

Golob TF, Recker WW (2004) A method for relating type of crash to traffic flow characteristics on urban freeways. Transportation Research Part A: Policy and Practice 38(1):53–80.

Gordon TJ, Kostyniuk LP, Green PE, Barnes MA, Blower DF, Bogard SE, Blankespoor AD, LeBlanc DJ, Cannon BR, McLaughlin SB (2013) A multivariate analysis of crash and naturalistic driving data in relation to highway factors.

Guido G, Vitale A, Astarita V, Saccomanno F, Giofré VP, Gallelli V (2012) Estimation of Safety Performance Measures from Smartphone Sensors. Procedia - Social and Behavioral Sciences 54:1095–1103.

Guo F, Fang Y (2013) Individual driver risk assessment using naturalistic driving data. Accident; analysis and prevention 61:3–9.

Guo F, Klauer SG, Hankey JM, Dingus TA (2010) Near Crashes as Crash Surrogate for Naturalistic Driving Studies. Transportation Research Record 2147(1):66–74.

Hallmark SL, Tyner S, Oneyear N, Carney C, McGehee D (2015) Evaluation of driving behavior on rural 2-lane curves using the SHRP 2 naturalistic driving study data. Journal of safety Research 54:17–27.

References

Hanowski RJ, Perez MA, Dingus TA (2005) Driver distraction in long-haul truck drivers. Transportation Research Part F: Traffic Psychology and Behaviour 8(6):441–458.

Hardy J, Campbell M (2013) Contingency Planning Over Probabilistic Obstacle Predictions for Autonomous Road Vehicles. IEEE Trans. Robot. 29(4):913–929.

Hassan HM, Abdel-Aty MA (2013) Predicting reduced visibility related crashes on freeways using real-time traffic flow data. Journal of safety Research 45:29–36.

Haversine formula - Calculate geographic distance on earth (2015). https://www.igismap.com/haversine-formula-calculate-geographic-distance-earth/

Hayward JC (1972) Near miss determination through use of a scale of danger. 51st Annual Meeting of the Highway Research Board Location: Washington District of Columbia, United States January, 17-21.

Herbrich R (2001) Learning kernel classifiers: theory and algorithms. MIT press.

Hörl S, Ciari F, Axhausen KW (2016) Recent perspectives on the impact of autonomous vehicles. Arbeitsberichte Verkehrs-und Raumplanung 1216.

Horst RVD (1991) Time-to-collision as a cue for decision-making in braking. Vision in Vehicles-III. North-Holland, Amsterdam.

Hossain M (2011) Development of a real-time proactive road safety management system for urban expressways. Doctoral Thesis. Tokyo Institute of Technology.

Hossain M, Muromachi Y (2012) A Bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. Accident; analysis and prevention 45:373–381.

Hossain M, Muromachi Y (2013) A real-time crash prediction model for the ramp vicinities of urban expressways. IATSS research 37(1):68–79.

Howard TM (2009) Adaptive model-predictive motion planning for navigation in complex environments. Doctoral Thesis. Carnegie Mellon University.

https://www.facebook.com/AnalyticsVidhya/ (2016) A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python).

https://www.facebook.com/AnalyticsVidhya/ (2018) Introduction to k-Nearest Neighbors: Simplified (with implementation in Python). https://www.analyticsvidhya.com/blog/2018/03/introduction-k-neighbours-algorithm-clustering/

https://www.facebook.com/pages/Statistics-Solutions What is Logistic Regression? - Statistics Solutions. https://www.statisticssolutions.com/what-is-logistic-regression/

Hubmann C, Becker M, Althoff D, Lenz D, Stiller C (2017) Decision making for autonomous driving considering interaction and uncertain prediction of surrounding vehicles. In: 2017 IEEE Intelligent Vehicles Symposium (IV 2017). Los Angeles, California, USA, 11-14 June 2017. IEEE, Piscataway, NJ, pp 1671–1678.

Hundelshausen F von, Himmelsbach M, Hecker F, Mueller A, Wuensche H-J (2008) Driving with tentacles: Integral structures for sensing and motion. J. Field Robotics 25(9):640–673.

Hupfer C (1997) Deceleration to safety time (DST)—a useful figure to evaluate traffic safety. In: ICTCT Conference Proceedings of Seminar, vol 3, pp 5–7.

Hydén C (1977) A Traffic-Conflicts Technique for Determining Risk. University of Lund.

References

Hydén C (1987) The development of a method for traffic safety evaluation: The Swedish Traffic Conflicts Technique. Bulletin Lund Institute of Technology, Department(70).

Ikeda H, Kaneko Y, Matsuo T, Tsuji K (1999) Abnormal incident detection system employing image processing technology. In: 1999 IEEE. IEEE, pp 748–752.

James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning, vol 112. Springer.

Jeon Jh, Cowlagi RV, Peters SC, Karaman S, Frazzoli E, Tsiotras P, Iagnemma K (2013) Optimal motion planning with the half-car dynamical model for autonomous high-speed driving. In: Staff I (ed) 2013 American Control Conference (ACC. IEEE, Washington, DC, USA, pp 188–193.

Jiménez F, Naranjo JE, Anaya JJ, García F, Ponz A, Armingol JM (2016) Advanced Driver Assistance System for Road Environments to Improve Safety and Efficiency. Transportation Research Procedia 14:2245–2254.

Jovanis PP, Shankar V, Aguero-Valverde J, Wu K-F, Greenstein A (2012) Analysis of Existing Data: Prospective Views on Methodological Paradigms. Transportation Research Board, Washington, D.C.

Kala R, Warwick K (2013) Motion planning of autonomous vehicles in a non-autonomous vehicle environment without speed lanes. Engineering Applications of Artificial Intelligence 26(5-6):1588–1601.

Kammel S, Ziegler J, Pitzer B, Werling M, Gindele T, Jagzent D, Schöder J, Thuy M, Goebl M, Hundelshausen F von, Pink O, Frese C, Stiller C (2009) Team AnnieWAY's Autonomous System for the DARPA Urban Challenge 2007. In: Buehler M, Iagnemma K, Singh S (eds) The DARPA urban challenge. Autonomous vehicles in city traffic / Martin Buehler, Karl Iagnemma, Sanjiv Singh (eds.), vol 56. Springer, Berlin, pp 359–391.

Katrakazas C (2017) Developing an advanced collision risk model for autonomous vehicles. Thesis (Ph.D.), Loughborough University.

Katrakazas C, Antoniou C, Yannis G (2019) Time Series Classification Using Imbalanced Learning for Real-Time Safety Assessment. Proceedings of the Transportation Research Board (TRB) 98th Annual Meeting, Washington, D.C., January 13-17.

Katrakazas C, Quddus M, Chen W-H, Deka L (2015) Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions. Transportation Research Part C: Emerging Technologies 60:416–442.

Kearney AT (2016) How automakers can survive the self-driving era. Retrieved November 14:2017.

Kiefer RJ, LeBlanc DJ, Flannagan CA (2005) Developing an inverse time-to-collision crash alert timing approach based on drivers' last-second braking and steering judgments. Accident; analysis and prevention 37(2):295–303.

Klauer SG, Dingus TA, Neale VL, Sudweeks J d., Ramsey DJ (2006) The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. National Highway Traffic Safety Administration. April.

Kuhnt F, Kohlhaas R, Schamm T, Zöllner JM (2015) Towards a unified traffic situation estimation model— Street-dependent behaviour and motion models. In: 2015 18th International Conference on Information Fusion (Fusion), pp 1223–1229.

References

Lambert (2016) Transcript: Elon Musk's press conference about Tesla Autopilot under v8.0 update [Part 1].

Laureshyn A, Johnsson C, Madsen T, Varhelyi A, Goede M de, Svensson Å, Saunier N, van Haperen W (2017) Exploration of a method to validate surrogate safety measures with a focus on vulnerable road users. In: p 11.

Laureshyn A, Svensson A, Hydén C (2010) Evaluation of traffic safety, based on micro-level behavioural data: theoretical framework and first implementation. Accident; analysis and prevention 42(6):1637–1646.

LeBlanc D (2006) Road departure crash warning system field operational test: methodology and results. volume 1: technical report.

Lee C, Abdel-Aty M, Hsia L (2006) Potential real-time indicators of sideswipe crashes on freeways. Transportation Research Record (1953):41–49.

Lee C, Hellinga B, Saccomanno F (2003) Real-Time Crash Prediction Model for Application to Crash Prevention in Freeway Traffic. Transportation Research Record 1840(1):67–77.

Lee U, Yoon S, Shim H, Vasseur P, Demonceaux C (2014) Local path planning in a complex environment for self-driving car. In: IEEE 4th Annual International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER), 2014. 4-7 June 2014, Hong Kong. IEEE, Piscataway, NJ, pp 445–450.

Lefevre S, Laugier C, Ibanez-Guzman J (2012) Evaluating risk at road intersections by detecting conflicting intentions. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012. 7-12 Oct. 2012, Vilamoura, Algarve, Portugal. IEEE, Piscataway, NJ, pp 4841–4846.

Lefèvre S, Vasquez D, Laugier C (2014) A survey on motion prediction and risk assessment for intelligent vehicles. Robomech J 1(1):658.

Li X, Lord D, Zhang Y, Xie Y (2008) Predicting motor vehicle crashes using support vector machine models. Accident Analysis & Prevention 40(4):1611–1618.

Li Y, Zheng Y, Wang J, Kodaka K, Li K (2018) Crash probability estimation via quantifying driver hazard perception. Accident; analysis and prevention 116:116–125.

Li Z, Liu P, Wang W, Xu C (2012) Using support vector machine models for crash injury severity analysis. Accident; analysis and prevention 45:478–486.

Lin L, Wang Q, Sadek AW (2015) A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. Transportation Research Part C: Emerging Technologies 55:444–459.

Litman T (2015) Autonomous vehicle implementation predictions: Implications for transport planning. Transportation Research Board Annual Meeting 2014.

Lord D (1996). Analysis of Pedestrian Conflicts with Left-Turning Traffic. Transportation Research Record 1538(1):61–67.

Lv Y, Tang S, Zhao H (2009) Real-Time Highway Traffic Accident Prediction Based on the k-Nearest Neighbor Method. In: 2009 International Conference on Measuring Technology and Mechatronics Automation. ICMTMA 2009, Zhangjiajie, Hunan, China, April 11-12, 2009. IEEE Computer Society Press, Los Alamitos, Calif., pp 547–550.

References

Maheshwari V (2019) Implementation of Random Forest – Data Driven Investor – Available online at https://medium.com/datadriveninvestor/implementation-of-random-forest-7bd04ceedeb2.

Martin A (2013) Interactive Motion Prediction using Game Theory. Master Thesis, University of Padova.

McNaughton M, Urmson C, Dolan JM, Lee J-W (2011) Motion planning for autonomous driving with a conformal spatiotemporal lattice. In: Bicchi A (ed) 2011 IEEE International Conference on Robotics and Automation. (ICRA); 9-13 May 2011, Shanghai, China. IEEE, Piscataway, NJ, pp 4889–4895.

Meinel HH (1998) Automotive millimeterwave radar history and present status. In: 1998 28th European Microwave Conference, vol 1, pp 619–629.

Michael PG, Leeming FC, Dwyer WO (2000) Headway on urban streets: observational data and an intervention to decrease tailgating. Transportation Research Part F: Traffic Psychology and Behaviour 3(2):55–64.

Mima H, Ikeda K, Shibata T, Fukaya N, Hitomi K, Bando T (2009) A rear-end collision warning system for drivers with support vector machines. In: 2009 IEEE/SP 15th Workshop on Statistical Signal Processing, pp 650–653.

Minderhoud MM, Bovy PHL (2001) Extended time-to-collision measures for road traffic safety assessment. Accident Analysis & Prevention 33(1):89–97.

Mohamed SA, Mohamed K, Al-Harthi. HA (2017) Investigating Factors Affecting the Occurrence and Severity of Rear-End Crashes. Transportation Research Procedia 25:2098–2107.

Molnar LJ, Eby DW, Bogard SE, LeBlanc DJ, Zakrajsek JS (2018) Using naturalistic driving data to better understand the driving exposure and patterns of older drivers. Traffic injury prevention 19(sup1): S83-S88.

Muronga K, Venter K (2014) Naturalistic driving data: managing and working with large databases for road and traffic management research. In: Proceedings of the 33rd Southern African Transport Conference (SATC 2014), vol 7, p 10.

Murphy KP (2012) Machine learning: a probabilistic perspective. MIT press.

Murphy L, Newman P (2011) Risky planning: Path planning over costmaps with a probabilistically bounded speed-accuracy tradeoff. In: Bicchi A (ed) 2011 IEEE International Conference on Robotics and Automation. (ICRA); 9-13 May 2011, Shanghai, China. IEEE, Piscataway, NJ, pp 3727–3732.

Murty MN, Raghava R (2016) Kernel-Based SVM. In: Murty MNa, Raghava R (eds) Support vector machines and perceptrons. Learning, optimization, classification, and application to social networks / M.N. Murty, Rashmi Raghava, vol 2. Springer, Switzerland, pp 57–67.

National Highway Traffic Safety Administration and U.S. Department of Transportation 2011 Data: Overview.

Neale VL, Dingus TA, Klauer SG, Sudweeks J, Goodman M (2005) An overview of the 100-car naturalistic study and findings. National Highway Traffic Safety Administration, Paper 5:400.

Oh C, Oh J-S, Ritchie S, Chang M (2001) Real-time estimation of freeway accident likelihood. In: 80th Annual Meeting of the Transportation Research Board, Washington, DC.

Ohta H (1993) Individual differences in driving distance headway. Vision in vehicles 4:91–100.

References

Ong SCW, Png SW, Hsu D, Lee WS (2010) Planning under Uncertainty for Robotic Tasks with Mixed Observability. The International Journal of Robotics Research 29(8):1053–1068.

Paden B, Cap M, Yong SZ, Yershov D, Frazzoli E (2016) A Survey of Motion Planning and Control Techniques for Self-Driving Urban Vehicles. IEEE Trans. Intell. Veh. 1(1):33–55.

Pande A, Abdel-Aty M (2006) Comprehensive Analysis of the Relationship between Real-Time Traffic Surveillance Data and Rear-End Crashes on Freeways. Transportation Research Record 1953(1):31–40.

Pande A, Abdel-Aty M (2007) Multiple-model framework for assessment of real-time crash risk. Transportation Research Record (2019):99–107.

Pande A, Das A, Abdel-Aty M, Hassan H (2011) Estimation of Real-Time Crash Risk. Transportation Research Record 2237(1):60–66.

Paone J, Bolme D, Ferrell R, Aykac D, Karnowski T (2015) Baseline face detection, head pose estimation, and coarse direction detection for facial data in the SHRP2 naturalistic driving study. In: IV2015. 2015 IEEE Intelligent Vehicles Symposium: June 28-July 1, 2015, COEX, Seoul, Korea. IEEE, [Piscataway, New Jersey], pp 174–179.

Patel S (2017) Chapter 2: SVM (Support Vector Machine) — Theory – Available online at https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, others (2011) Scikit-learn: Machine Learning in Python Journal of Machine Learning Research.

Perkins SR, Harris J (1968) Traffic conflict characteristics-accident potential at intersections. Highway Research Record (225).

Pivtoraiko M, Kelly A (2005) Efficient constrained path planning via search in state lattices. In: International Symposium on Artificial Intelligence, Robotics, and Automation in Space, pp 1–7.

Qu X, Wang W, Wang W, Liu P (2013) Real-time freeway sideswipe crash prediction by support vector machine. IET Intelligent Transport Systems 7(4):445–453.

Regan M, Williamson A, Grzebieta R, Charlton J, Lenne, Watson B, Haworth N, Rakotonirainy A, Woolley, Anderson, Senserrick T, Young K (2013) The Australian 400-Car Naturalistic Driving Study: Innovation in Road Safety Research and Policy. In: Proceedings of the 2013 Australasian road safety research, policing & education conference, Brisbane, Queensland.

Riener A, Ferscha A (2009) Effect of Proactive Braking on Traffic Flow and Road Throughput. In: Turner SJ (ed) DS-RT 2009. Proceedings 13th IEEEACM International Symposium on Distributed Simulation and Real-Time Applications 25-28 October 2009, Singapore. IEEE Computer Society Press, Los Alamitos Calif., pp 157–164.

Roshandel S, Zheng Z, Washington S (2015) Impact of real-time traffic characteristics on freeway crash occurrence: systematic review and meta-analysis. Accident; analysis and prevention 79:198–211.

Sargent DJ (2001) Comparison of artificial neural networks with other statistical approaches: results from medical data sets. Cancer 91(8 Suppl):1636–1642.

References

Saunier N, Sayed T, Ismail K (2010) Large-Scale Automated Analysis of Vehicle Interactions and Collisions. Transportation Research Record 2147(1):42–50.

Schroder J, Gindele T, Jagszent D, Dillmann R (2008) Path planning for cognitive vehicles using risk maps. In: 2008 IEEE Intelligent Vehicles Symposium. IEEE Xplore, Piscataway, N.J., pp 1119–1124.

Schwarting W, Alonso-Mora J, Rus D (2018) Planning and Decision-Making for Autonomous Vehicles. Annu. Rev. Control Robot. Auton. Syst. 1(1):187–210.

Seaman S, Lee J, Angell L, Mehler B, Seppelt B, Reimer B (2016) Exploring Generalizability of Field Experiment Radio Tasks with Naturalistic Driving Data. In: Green P, Pfleging B, Kun AL, Liang Y, Meschtscherjakov A, Fröhlich P (eds) AutomotiveUI 2016. 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: adjunct proceedings: October 24-26, 2016, Ann Arbor, MI, USA. Association for Computing Machinery, New York, New York, pp 111–116.

Seshadri K, Juefei-Xu F, Pal DK, Savvides M, Thor CP (2015) Driver cell phone usage detection on Strategic Highway Research Program (SHRP2) face view videos. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on. Date, 7-12 June 2015. IEEE, Piscataway, New Jersey, pp 35–43.

Shankar V, Jovanis PP, Aguero-Valverde J, Gross F (2008) Analysis of Naturalistic Driving Data. Transportation Research Record 2061(1):1–8.

Shbeeb L (2000) Development of traffic conflicts technique for different environments: A comparative study of pedestrian conflicts in Sweden and Jordan, Lund University.

Singh S (2003) Driver attributes and rear-end crash involvement propensity. National Highway Traffic Safety Administration. March.

Singh S (2015) Critical reasons for crashes investigated in the national motor vehicle crash causation survey. National Highway Traffic Safety Administration. February 1–2.

Sivak M, Schoettle B (2015) Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles. University of Michigan Transportation Research Institute.

Sobhani A, Young W, Bahrololoom S, Sarvi M, others (2013) Calculating time-to-collision for analysing right turning behaviour at signalised intersections. Road & Transport Research: A Journal of Australian and New Zealand Research and Practice 22(3):49.

Song W, Xiong G, Chen H (2016) Intention-Aware Autonomous Driving Decision-Making in an Uncontrolled Intersection. Mathematical Problems in Engineering 2016.

Songchitruksa P, Tarko A (2006) Practical Method for Estimating Frequency of Right-Angle Collisions at Traffic Signals. Transportation Research Record 1953:89–97.

Staubach M (2009) Factors correlated with traffic accidents as a basis for evaluating Advanced Driver Assistance Systems. Accident; analysis and prevention 41(5):1025–1033.

Stewart B (1998) Improving performance of naive bayes classifier by including hidden variables. In: Pobil APd, Mira J, Ali M (eds) 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems, IEA-98-AIE, Benicàssim, Castellón, Spain, June 1-4, 1998, proceedings. Vol.1, Methodology and tools in knowledge-based systems, vol 1415. Springer, Berlin, London, pp 272–280.

References

Stilgoe J (2017) Machine Learning, Social Learning and the Governance of Self-Driving Cars. SSRN Journal.

Stutts J, Feaganes J, Reinfurt D, Rodgman E, Hamlett C, Gish K, Staplin L (2005) Driver's exposure to distractions in their natural driving environment. Accident; analysis and prevention 37(6):1093–1101.

Sun J, Sun J (2015) A dynamic Bayesian network model for real-time crash prediction using traffic speed conditions data. Transportation Research Part C: Emerging Technologies 54:176–186.

Tabachnick BG, Fidell LS (2007) Using multivariate statistics, Pearson International ed., 5th ed. Pearson/A&B, Boston Mass., London.

Tarko A, Romero MA, Bandaru VK, Ariyur KB, Lizarazo CG (2017) Feasibility of Tracking Vehicles at Intersections with a Low-end LiDAR. In: Proceedings of the 96th Annual Meeting of the Transportation Research Board. Washington, DC. January 8-12.

Tarko AP, Songchitruksa P (2005) Estimating the frequency of crashes as extreme traffic events. In: 84th Annual Meeting of the Transportation Research Board. Washington, DC. January 9-13.

Thomopoulos N, Givoni M (2015) The autonomous car—a blessing or a curse for the future of low carbon mobility? An exploration of likely vs. desirable outcomes. Eur J Futures Res 3(1):31.

Treat, Tumbas NS, McDonald ST, Shinar D, Hume RD (1979) Tri-level study of the causes of traffic accidents executive summary, Technical Report.

Triggs TJ, Harris WG (1982) Reaction time of drivers to road stimuli, Human Factors Report No. HFR-12, Monash University, Australia.

UDRIVE. European Naturalistic Driving Study. http://www.udrive.eu/

van der Horst ARA (1990) A time-based analysis of road user behaviour in normal and critical encounters. Doctoral Thesis. Delft University of Technology, The Netherlands.

van der Horst R, Hogema J (1993) Time-to-collision and collision avoidance systems. Proceedings of the 6th ICTCT Workshop 1993: Safety Evaluation of Traffic Systems: Traffic Conflicts and Other Measures, 109–121.

van Winsum W, Brookhuis KA, Waard D de (2000) A comparison of different ways to approximate time-to-line crossing (TLC) during car driving. Accident Analysis & Prevention 32(1):47–56.

Varaiya P (1993) Smart cars on smart roads: problems of control. IEEE Trans. Automat. Contr. 38(2):195–207.

Várhelyi A (1998) Drivers' speed behaviour at a zebra crossing: a case study. Accident Analysis & Prevention 30(6):731–743.

Venter K (2014) An exploratory study into South African novice driver behaviour. Master Thesis. Stellenbosch University.

Verschuur WLG, Hurts K (2008) Modeling safe and unsafe driving behaviour. Accident; analysis and prevention 40(2):644–656.

Vlahogianni EI, Barmpounakis EN (2017) Driving analytics using smartphones: Algorithms, comparisons and challenges. Transportation Research Part C: Emerging Technologies 79:196–206.

Vogel K (2003) A comparison of headway and time to collision as safety indicators. Accident Analysis & Prevention 35(3):427–433.

References

Waggenaar WA, Reason JT (1990) Types and tokens in road accident causation. Ergonomics 33(10-11):1365–1375.

Wåhlberg AE af (2003) Some methodological deficiencies in studies on traffic accident predictors. Accident Analysis & Prevention 35(4):473–486.

Wang J, Zhou H (2018) Using Naturalistic Driving Study Data to Evaluate the Effects of Intersection Balance on Driver Behavior at Partial Cloverleaf Interchange Terminals. Transportation Research Record 2672(38):255–265.

Wang W, Qu X, Wang W, Liu P (2013) Real-time freeway sideswipe crash prediction by support vector machine. IET Intelligent Transport Systems 7(4):445–453.

Wang Y, Ieda H, Mannering F (2003) Estimating rear-end accident probabilities at signalized intersections: occurrence-mechanism approach. Journal of Transportation engineering 129(4):377–384.

Ward J, Agamennoni G, Worrall S, Nebot E (2014a) Vehicle collision probability calculation for general traffic scenarios under uncertainty. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings, pp 986–992.

Ward J, Worrall S, Agamennoni G, Nebot E (2014b) The Warrigal Dataset: Multi-Vehicle Trajectories and V2V Communications. IEEE Intell. Transport. Syst. Mag. 6(3):109–117.

Wei J, Snider JM, Gu T, Dolan JM, Litkouhi B (2014) A behavioral planning framework for autonomous driving. In: 2014 IEEE Intelligent Vehicles Symposium Proceedings, pp 458–464.

Wei J, Snider JM, Kim J, Dolan JM, Rajkumar R, Litkouhi B (2013) Towards a viable autonomous driving research platform. In: 2013 IEEE Intelligent Vehicles Symposium (IV), pp 763–770.

White CC, White DJ (1989) Markov decision processes. European Journal of Operational Research 39(1):1–16.

Wille JM, Saust F, Maurer M (2010) Comprehensive treated sections in a trajectory planner for realizing autonomous driving in Braunschweig'S urban traffic. In: Staff I (ed) 2010 13th International IEEE Conference on Intelligent Transportation Systems. IEEE, Funchal, Portugal. pp 647–652.

Worrall S, Agamennoni G, Nieto J, Nebot E (2012) A Context-Based Approach to Vehicle Behavior Prediction. IEEE Intell. Transport. Syst. Mag. 4(3):32–44.

Wu K-F, Aguero-Valverde J, Jovanis PP (2014) Using naturalistic driving data to explore the association between traffic safety-related events and crash risk at driver level. Accident; analysis and prevention 72:210–218.

Wu K-F, Jovanis PP (2012) Crashes and crash-surrogate events: exploratory modeling with naturalistic driving data. Accident; analysis and prevention 45:507–516.

Xiong X, Chen L, Liang J (2018) Vehicle Driving Risk Prediction Based on Markov Chain Model. Discrete Dynamics in Nature and Society 2018(8):1–12.

Xu C, Liu P, Wang W (2016) Evaluation of the predictability of real-time crash risk models. Accident; analysis and prevention 94:207–215.

Xu C, Wang W, Liu P (2013) A Genetic Programming Model for Real-Time Crash Prediction on Freeways. IEEE Trans. Intell. Transport. Syst. 14(2):574–586.

References

Xu C, Wang W, Liu P, Li Z (2015) Calibration of crash risk models on freeways with limited real-time traffic data using Bayesian meta-analysis and Bayesian inference approach. Accident; analysis and prevention 85:207–218.

Xu W, Pan J, Wei J, Dolan JM (2014) Motion planning under uncertainty for on-road autonomous driving. In: Robotics and Automation (ICRA), 2014 IEEE International Conference on. May 31 2014-June 7 2014: Hong Kong, China. IEEE. Piscataway, NJ. pp 2507–2512.

Yang H (2012) Simulation-based evaluation of traffic safety performance using surrogate safety measures. Doctoral Thesis. Rutgers University-Graduate School-New Brunswick.

Yu R, Abdel-Aty M (2013) Utilizing support vector machine in real-time crash risk evaluation. Accident; analysis and prevention 51:252–259.

Yu R, Abdel-Aty M, Ahmed M (2013) Bayesian random effect models incorporating real-time weather and traffic data to investigate mountainous freeway hazardous factors. Accident Analysis & Prevention 50:371–376.

Zhang S, Deng W, Zhao Q, Sun H, Litkouhi B (2013) Dynamic Trajectory Planning for Vehicle Autonomous Driving. In: Proceedings: 2013 IEEE International Conference on Systems, Man and Cybernetics. SMC 2013: 13-16 October 2013: Manchester, United Kingdom. IEEE Computer Society, Los Alamitos, CA, pp 4161–4166.

Zhao P, Chen J, Mei T, Liang H (2011) Dynamic motion planning for autonomous vehicle in unknown environments. In: 2011 IEEE Intelligent Vehicles Symposium (IV 2011). Germany 5-9 June 2011. IEEE, Piscataway, NJ, pp 284–289.

Zheng Y, Wang J, Li X, Yu C, Kodaka K, Li K (2014) Driving risk assessment using cluster analysis based on naturalistic driving data. In: 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp 2584–2589.

Zheng Z, Ahn S, Monsere CM (2010) Impact of traffic oscillations on freeway crash occurrences. Accident; analysis and prevention 42(2):626–636.

Ziegler J, Bender P, Schreiber M, Lategahn H, Strauss T, Stiller C, Dang T, Franke U, Appenrodt N, Keller CG, Kaus E, Herrtwich RG, Rabe C, Pfeiffer D, Lindner F, Stein F, Erbs F, Enzweiler M, Knoppel C, Hipp J, Haueis M, Trepte M, Brenk C, Tamke A, Ghanaat M, Braun M, Joos A, Fritz H, Mock H, Hein M, Zeeb E (2014) Making Bertha Drive—An Autonomous Journey on a Historic Route. IEEE Intell. Transport. Syst. Mag. 6(2):8–20.

Ziegler J, Stiller C (2009) Spatiotemporal state lattices for fast trajectory planning in dynamic on-road driving scenarios. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St Louis, Missouri, 10-15 October 2009. IEEE, Piscataway, NJ, pp 1879–1884.

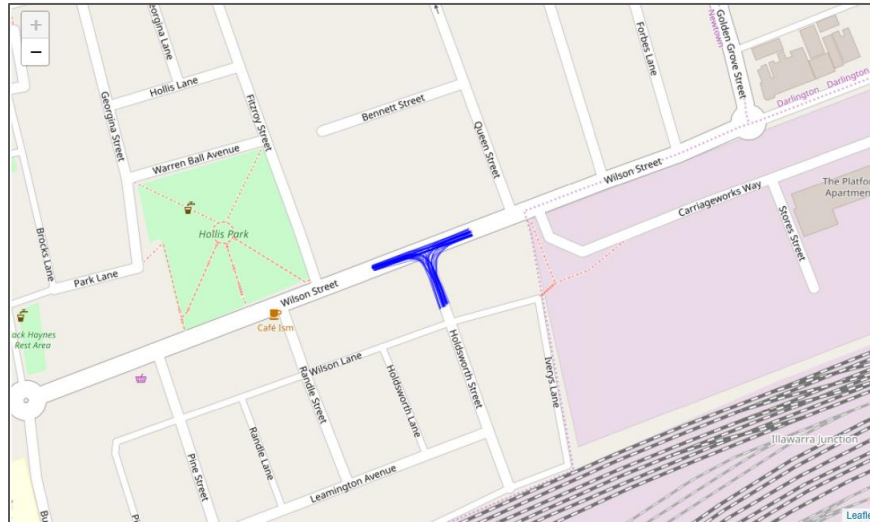# Appendix A: Trajectories in the Bounding Area for Driver 1 and 2

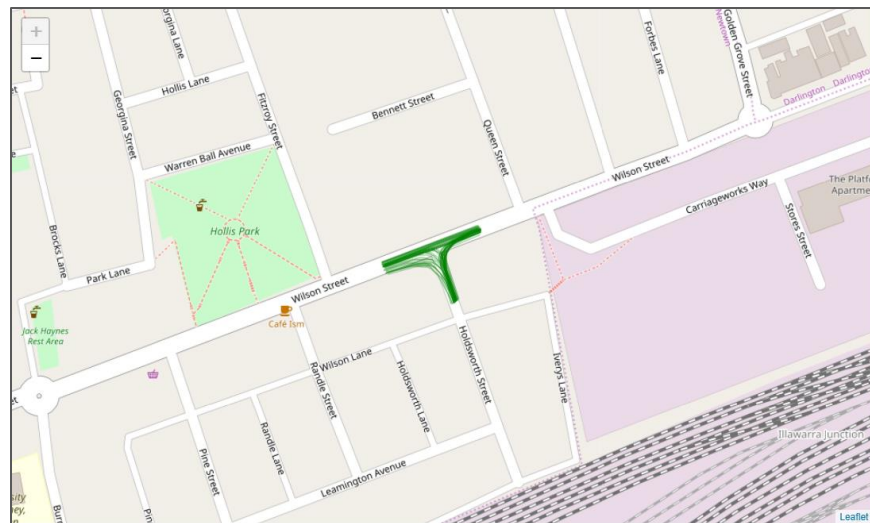**Figure A.1** Trajectories in the bounding area for driver 1

Figure A.2 Trajectories in the bounding area for driver 2

Declaration concerning the Master's Thesis

I hereby confirm that the presented thesis work has been done independently and the materials and methods used and quoted has been properly referenced and acknowledged. This thesis has not previously been published or submitted elsewhere for purposes of assessment.

_____

Place and date   Signature