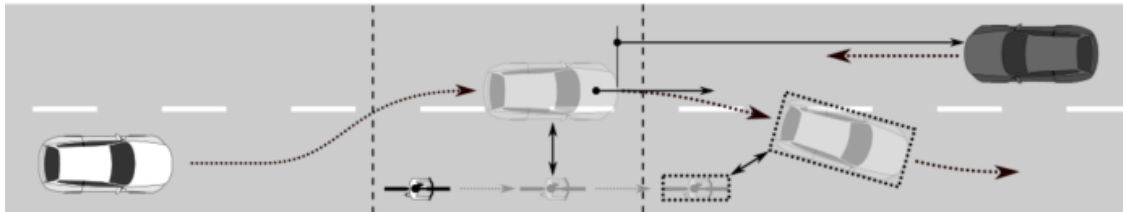




CHALMERS
UNIVERSITY OF TECHNOLOGY



Detecting and Evaluating Risky Behaviors in Overtaking Scenarios for Safer Driving

Master's Thesis in MPSYS and MPALG

Junzhao Cheng
Yanchen Lin

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

MASTER'S THESIS IN MPSYS AND MPALG

Detecting and Evaluating Risky Behaviors in Overtaking Scenarios for Safer Driving

Junzhao Cheng
Yanchen Lin



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Mechanics and Maritime Sciences
Division of Active Safety
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025

Detecting and Evaluating Risky Behaviors in Overtaking Scenarios for Safer Driving
Junzhao Cheng
Yanchen Lin

© Junzhao Cheng, Yanchen Lin, 2025.

Supervisor: Andrew Backhouse and Carina Björnsson, Volvo Cars Corporation
Examiner: Marco Dozza, Department of Mechanics and Maritime Sciences

Master's Thesis 2025
Department of Mechanics and Maritime Sciences
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 31 772 1000

Cover: An ideal overtaking cyclist scenario with oncoming vehicle

Typeset in L^AT_EX
Gothenburg, Sweden 2025

Detecting and Evaluating Risky Behaviors in Overtaking Scenarios for Safer Driving

Junzhao Cheng

Department of Electric Engineering
Chalmers University of Technology

Yanchen Lin

Department of Computer Science and Engineering
Chalmers University of Technology

Abstract

While cycling has become a popular way of modern transportation, concerns about cyclist safety remain significant. One typical and potentially risky scenario arises when motor vehicles pass (including overtake) cyclists, which may lead to close passing incidents and endanger the vulnerable road users.

Previous works intensively investigated a typical scenario where an ego vehicle is overtaking a cyclist on a two-lane rural road with possible oncoming vehicles present on the adjacent lane. Experiments have been conducted to study the characteristics in vehicle-cyclist interaction during overtaking. Many studies segment the overtaking period into phases to better model the maneuver. There are also traffic laws that requires a minimum lateral distance when a motor vehicle passes cyclists.

In this thesis, we used real-world data from Volvo Cars Corporation to detect and evaluate risky passing and overtaking behavior. We used and further developed a tool that extracts scenarios when ego vehicles pass cyclist. Then output information derived from signals in these scenarios. With that information, we built a rule-based model and applied it to different phases of the overtaking process. The model used four safety metrics: Variable-Adjusted Minimum Passing Distance, Perceived Risk Score, Minimum Distance Returning, and the Lateral-Time Risk Index. We conducted statistical analysis on the data and the model output. We also verified some of the results by manually reviewing the scenario on a visualization system to evaluate the correctness of our model.

Furthermore, we discussed some special or difficult cases, such as overtaking a cyclist group, sensor unreliability at night, overtaking maneuvers happening on curved roads. We also pointed out the possible limitations in the data source, sensor fusion process and other factors. Lastly, we showed our perspective on the future works that can be based on our thesis work.

The work presented in this thesis can help better understand how drivers overtake cyclists, and it may be useful for providing post-event feedback to enhance driving safety, or provide insights to user-based insurance systems.

Keywords: Overtaking, Risky Driving Detection, Scenario Extraction, Cyclist Safety, Safety Metrics, Advanced Driver Assistance System (ADAS), User-Based Insurance, Post-Event Feedback.

Preface

This report presents the outcome of our master's thesis project carried out at the Department of Mechanics and Maritime Sciences at Chalmers University of Technology and Volvo Cars Corporation during the spring of 2025.

Acknowledgements

We are very grateful to Volvo Car Corporation for its support in helping us complete our graduation project. In addition, we would like to thank Marco, Andy and Carina for their patient guidance and companionship. Every meeting is a brainstorming session and we enjoy every discussion. We would also like to thank our colleagues in the collision avoidance team for their care and guidance and the ADAM team for their great help in our scene capture. We really enjoy working in the Volvo office every day, which is full of cooperation, enthusiasm and discussion. We also like the weekly fika time.

Junzhao Cheng, Yanchen Lin, Gothenburg, May 2025

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

ADAS	Advanced Driver Assistance Systems
AV	Autonomous Vehicle
ED	Euclidean Distance
FT	Field Test
GPS	Global Positioning System
HDV	Human-Driven Vehicle
LC	Lateral Clearance
LD	Lateral Distance
LIDAR	Light Detection and Ranging
LTRI	LD-TTD Risk Index
MPD	Minimum Passing Distance
MDR	Minimum Distance Returning
NDS	Naturalistic Driving Study
PRS	Perceived Risk Score
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
TTC	Time To Collision
TTC_{cyc}	Time To Collision with Cyclists
TTC_{on}	Time To Collision with Oncoming Vehicle
TTD_{cyc}	Time To Danger with Cyclists
V_{ego}	Speed of the ego vehicle
VAMPD	Variable-Adjusted Minimum Passing Distance
VCC	Volvo Cars Corporation
VRU	Vulnerable Road User

Contents

List of Acronyms	ix
Nomenclature	xi
List of Figures	xv
List of Tables	xvii
1 Introduction	1
1.1 Background	1
1.2 Literature Review	1
1.2.1 Terminology Unification	2
1.2.1.1 Ego Vehicle	2
1.2.1.2 Lateral Distance	2
1.2.1.3 Ego Velocity	2
1.2.2 Overtaking Cyclist on Two-lane Rural Roads	2
1.2.3 Experiments	3
1.2.4 Perceived Safety and Driver Behavior	4
1.2.5 Overtaking Phases	4
1.2.6 ADAS Support	4
1.2.7 Passing Distance	5
1.3 Purpose	5
1.4 Goals	6
2 Theory	7
2.1 Defining Overtaking Maneuvers	7
2.2 Classification of Overtaking Strategies	7
2.3 Safety Metrics Introduction	8
2.3.1 Variable-Adjusted Minimum Passing Distance (VAMPD)	9
2.3.1.1 Advantages and Limitations	9
2.3.1.1.1 Advantages	9
2.3.1.1.2 Limitations	9
2.3.2 Perceived Risk Score	10
2.3.2.1 Advantages and Limitations	10
2.3.2.1.1 Advantages	10
2.3.2.1.2 Limitations	10
2.3.3 LD-TTD Risk Index(LTRI)	11

2.3.3.1	Advantages and Limitations	12
2.3.3.1.1	Advantages	12
2.3.3.1.2	Limitations	12
2.3.4	Minimum Distance Returning (MDR)	12
2.3.4.1	Advantages and Limitations	12
2.3.4.1.1	Advantages	12
2.3.4.1.2	Limitations	12
2.4	Scenarios Classification and Scope	12
2.5	Phase Definitions and Model Applications	13
2.5.1	Drive-by Scenario Modeling	13
2.5.1.1	Approaching Phase	13
2.5.1.2	Passing Phase	13
2.5.1.3	Departing Phase	14
2.5.2	Overtaking Scenario Modeling	14
2.5.2.1	Approaching Phase	14
2.5.2.2	Steering Away Phase	14
2.5.2.3	Passing Phase	15
2.5.2.4	Steering Back Phase	15
3	Methods	17
3.1	Data Description	17
3.2	Scenario Extractor	17
3.2.1	Finder	18
3.2.2	Phases Extraction	19
3.2.3	Output	19
3.2.4	Small Batch Run	20
3.2.5	Full-Scale Run	20
3.3	Post Filtering and Scenario Classification	20
3.4	Application of Safety Metrics	22
3.4.1	Variable-Adjusted Minimum Passing Distance (VAMPD) Im- plementation	22
3.4.1.1	Computation	22
3.4.1.2	Binary Safety Evaluation Criterion	22
3.4.2	Perceived Risk Score Implementation	22
3.4.2.1	Model Parameterization	23
3.4.2.2	Scoring Method Modification	24
3.4.2.3	Safety Evaluation Criterion	24
3.4.3	LD-TTD Risk Index(LTRI) Implementation	24
3.4.3.1	Safety Evaluation Criterion	25
3.4.4	Minimum Distance Returning (MDR)	25
3.4.4.1	Safety Evaluation Criterion	25
4	Results	27
4.1	Results on a Small Batch	27
4.1.1	Data Description	27
4.1.2	Comparison on Perceived Risk Score Model Before and After Scoring Method Modification	27

4.1.3	Drive-by Scenarios	29
4.1.3.1	LTRI	29
4.1.3.2	VAMPD	30
4.1.4	Overtaking Scenarios	30
4.1.4.1	LTRI	30
4.1.4.2	VAMPD	30
4.1.4.3	PRS	30
4.1.4.4	MDR	32
4.2	Results on Full Dataset	32
4.2.1	Data Description	32
4.2.2	drive-by Scenarios	33
4.2.2.1	LTRI	33
4.2.2.2	VAMPD	33
4.2.3	Overtaking Scenarios	34
4.2.3.1	LTRI	34
4.2.3.2	VAMPD	34
4.2.3.3	PRS	34
4.2.3.4	MDR	36
5	Discussion	37
5.1	Results from this thesis	37
5.1.1	Occurrence of Passing Types in Real World	37
5.1.2	Validation from Manual Inspection	37
5.1.3	Scoring Methods in Perceived Risk Score Model	38
5.1.4	Results on Safety Metrics	38
5.2	Special Cases	39
5.2.1	Overtaking a Group of Cyclists	39
5.2.2	Misclassification of Objects	39
5.2.3	Overtaking on Curved Road	40
5.3	Limitations	40
5.3.1	Rule-Based Safety Model	40
5.3.2	Data Source and Reproducibility	41
5.3.3	Sensor Fusion	42
5.3.3.1	Cyclist Dimension	42
5.3.3.2	Uncommon Type of Road User	43
5.3.4	Other Factors	43
5.3.4.1	Lack on Environmental Information	43
5.3.4.2	Cyclist Unsafe Behavior	43
5.4	Future Work	44
5.4.1	Scenario Extractor	44
5.4.1.1	Steering Maneuver	44
5.4.1.2	Overtaking Other Road Users	44
5.4.2	Models	45
5.4.2.1	Squeezing Strategy	45
5.4.2.2	Cyclist Velocity	45
5.4.2.3	Type of Oncoming Vehicle	46

6 Conclusion	47
Bibliography	49
A Appendix 1	I
B Appendix 2	III
C Appendix 3	V

List of Figures

2.1	Overtaking scenario definition for an exemplary flying maneuver, including overtaking phases and independent variables. LC is the lateral clearance between the ego vehicle and the cyclist, and V_{ego} is the speed of the ego vehicle at the passing moment. TTC_{on} is the time-to-collision between ego and oncoming vehicles at the passing moment. Source: Rasch et al. [41], *Drivers' and cyclists' safety perceptions in overtaking maneuvers*, 2022.	7
2.2	Two strategies of overtaking. The blue line indicates a flying maneuver that the ego vehicle (white) steers to overtake the cyclist before oncoming vehicle (black) comes close. The red line indicates a accelerative maneuver that the ego vehicle decelerates to let oncoming vehicle pass before overtaking the cyclist. Source: Rasch and Dozza [39], *Modeling Drivers' Strategy When Overtaking Cyclists in the Presence of Oncoming Traffic*.	8
2.3	Passing scenario and related safety metrics:(1)LD-TTD Risk Index (LTRI) applied in approaching phase,(2)Variable-Adjusted Minimum Passing Distance (VAMPD) applied in passing phase	13
2.4	Phases and metrics in overtaking scenario with multiple oncoming vehicles present. Since multiple oncoming vehicles may appear throughout different phases of the overtaking scenario (e.g., yellow, red, and blue vehicles), the Time to Collision (TTC) with the respective oncoming vehicle should be evaluated continuously as each vehicle comes into view. Overtaking scenario and related safety metrics: (1)LD-TTD Risk Index (LTRI) applied in approaching phase and steering away phase,(2)Perceived Risk Score (PRS) and Variable-Adjusted Minimum Passing Distance (VAMPD) applied in passing phase,(3)Minimum Distance Returning (MDR) applied in steering back phase	14
3.1	Data processing flow in the scope of this thesis work.	18
3.2	The figure shows the definition of drive-by scenarios, including phases and independent variables. LD is the lateral distance between the ego vehicle and the cyclist, and V_{ego} is the speed of the ego vehicle. TTD_{cyc} is the time-to-danger between ego vehicle and cyclist before passing.	21

3.3	The figure shows the definition of overtaking scenarios, including phases and independent variables. LD is the lateral distance between the ego vehicle and the cyclist, and V_{ego} is the speed of the ego vehicle. TTC_{on} is the time-to-collision between ego and oncoming vehicles at the passing moment. TTD_{cyc} is the time-to-danger between ego vehicle and cyclist before passing. ED is the Euclidean distance between ego vehicle and cyclist at the steering back phase.	21
4.1	Comparison of predicted cyclist perceived risk scores from the original scoring method and the altered weighted scoring methods on a 44-sample batch. The altered weighted model shows fewer low and high scores, with more mid-range predictions.	28
4.2	Comparison of original and weighted cyclist scoring on small batch data w.r.t ego longitudinal velocity and side-to-side lateral distance. .	29
4.3	Comparison of score differences between the weighted and original models on a small batch dataset. The histogram (left) shows the distribution of score differences, while the scatter plot (right) relates these differences to ego velocity and lateral distance.	29
4.4	Score Distribution for the perceived risk score metric on the small batch dataset. Using original scoring method for the driver model and weighted scoring method for the cyclist model.	31
4.5	Distribution of Minimum Distance Returning (MDR) values calculated across 44 overtaking maneuvers, indicating how closely the vehicle merges back after passing the cyclist.	32
4.6	Results of Perceived Risk Score distribution on full-scale overtaking dataset.	35
4.7	Results of Perceived Risk Score in the full-scale overtaking dataset using the weighted method, with respect to ego velocity and TTC with oncoming vehicle (0 if no oncoming vehicles) for the driver score, and with respect to ego velocity and side-to-side lateral distance for the cyclist score.	36
4.8	Distribution of Minimum Distance Returning values calculated across 1928 overtaking maneuvers.	36

List of Tables

2.1	Descriptions of safety-related models and metrics	9
2.3	Risk classification logic of the LD-TTD Risk Index (LTRI)[19]	11
3.2	Variables and annotations for overtaking scenarios	23
3.3	Median values of model parameters for driver and cyclist perceived safety models, adapted from [41]. These values were used directly in our implementation.	23
4.1	Summary of variables in the overtaking dataset (N = 44). All continuous variables are shown as mean (SD); all categorical variables as number of samples per level (percentage).	31
4.3	Summary of VAMPD results on drive-by scenarios in full dataset. All variables are shown as percentage.	33
4.5	Summary of VAMPD results on overtaking scenarios in the full dataset. All variables are shown as percentage.	34
4.7	Summary of variables in the full-scale overtaking dataset. All continuous variables are shown as mean (SD); all categorical variables as percentage.	35
A.1	Key variables and their definitions used in the study	I
B.1	Minimum Passing Distance (MPD) laws for overtaking cyclists in different countries and regions	III
C.1	Output variables from <i>EgoPassingCyclistFinder</i> in Scenario Extractor – General Information	V
C.3	Output Variables – Approaching & Steering Away Phase (Pre Passing Phase)	VI

1

Introduction

1.1 Background

Road traffic injury is a major global issue. According to the UN Global Road Safety Week 2023 report[36], approximately 1.19 million people die each year as a result of road traffic crashes, with many incidents caused by unsafe driver behaviors. More than half of all road traffic deaths are among vulnerable road users, including pedestrians, cyclists and motorcyclists. Some accidents happen because drivers are distracted, while others result from taking unnecessary risks. Overtaking is a risky maneuver and it plays a crucial role in many collision situations as it often results in severe or even deadly consequences due to high impact speed[35][24]. Vulnerable road users like cyclists face even higher risks in such situations.

Meanwhile, there is a growing global trend toward promoting cycling as a sustainable and healthy way of transportation[8]. However, in many regions, the road infrastructure like dedicated cycling lanes are often limited or entirely absent, forcing cyclists to share the same lanes with motor vehicles[25]. According to a WHO report[34] published on 2018, 86% of the cyclists are traveling with the absence of dedicated cycling infrastructures. This significantly increases the risk of crashes, especially during overtaking maneuvers. In such scenarios where motor vehicles interact with cyclists, ensuring safe interactions becomes critical.

Therefore, this thesis aims to investigate the characteristics of vehicle-cyclist interactions during overtaking, with a focus on modeling and evaluating risky driver behaviors in such scenarios, in order to promote safer driving.

1.2 Literature Review

There has been extensive research investigating scenarios where a motor vehicle is driving by or overtaking a cyclist (approach from behind and pass, with possible steering maneuver) , as these scenarios are common in the real world and can have lethal impacts on the cyclist and sometimes severe damage on the motor vehicle driver as well.

In this section, we begin by clarifying and unifying the terminology adopted in this thesis. We then provide a structured review of relevant literature regarding passing and overtaking scenarios, highlighting key findings and methods. Finally, we identify limitations in the existing research and outline potential directions for improvement and innovation, which motivate the contributions of this thesis.

1.2.1 Terminology Unification

Before summarizing the previous works, we want to first clarify on the terminology used in our thesis, as previous works used different terminology on the essentially identical concept. To address issues of redundancy and inconsistency, we choose the best set of terminology throughout this thesis. All related concepts will be referred to using the terms as defined in this section, regardless of the original variations used in previous works. This approach improves clarity and ensures consistent comparisons, while also reducing potential confusion for readers.

1.2.1.1 Ego Vehicle

In this thesis, we use the term *ego vehicle*, which is widely adopted in both academia and the vehicle industry, to refer to the vehicle under direct observation, modeling, or control.

1.2.1.2 Lateral Distance

Many previous research or regulations involve the side-to-side space to the cyclist when ego vehicle is passing it. While this concept is commonly referred to as *lateral distance* in some research papers, other studies use the term *lateral clearance*. On the contrary, some traffic laws or regulations refer to it as *passing distance*.

While all three terms are correct and scientific, in this thesis, we choose to adopt the term *lateral distance* because it is straight-forward, applicable outside the passing period, and aligns with the variable names used in an existing tool we used.

1.2.1.3 Ego Velocity

Similarly, previous works use different terms for the longitudinal velocity of the ego vehicle, most notably *ego speed* and *ego velocity*.

While both terms are widely accepted and straight-forward, in this thesis, we adopt the term *ego velocity* mostly because it aligns with the variable names used in the existing tool we used.

1.2.2 Overtaking Cyclist on Two-lane Rural Roads

There are many researchers investigating the factors affecting safety when overtaking cyclists, especially on two-lane rural roads with one possible oncoming vehicle in the neighboring lane[28][27][39][40].

Rural roads accommodate significant bicycle traffic volumes, has higher vehicle speed limit, and they usually are not equipped with cycling infrastructure, resulting in higher injury rate than the urban roads[27].

This has become a typical setting to investigate the safety in overtaking maneuvers for several reasons. First, the two-lane rural road setting means that the traffic environment is relatively simple, which reduces confounding variables and better enables controlled analysis of overtaking behavior. Second, rural roads are often narrower than urban roads, leaving limited lateral distance for vehicles to safely pass cyclists without steering. Third, such roads typically lack dedicated cycling

infrastructure, making it common for cyclists to ride directly within the travel lane. As a result, motor vehicles are frequently required to cross into the opposing lane to overtake, which is a potentially dangerous maneuver, especially in the presence of oncoming traffic.

1.2.3 Experiments

There have been many previous experiments designed to investigate how drivers overtake cyclists.

Some test-track experiments[30][20] use an instructed vehicle (ego vehicle) equipped with sensors that record vehicle signals like velocity and acceleration, information about traffic objects such as distance and velocity, and possibly record footage from cameras to further analyze the overtaking scenarios.

In addition to controlled experiments, some research record the real traffic on a fixed road and conduct a comprehensive analysis on the data, to provide insights on the passing distances and vehicle velocity in relation with road condition and the cyclist demographic characteristics.

In field-track experiments[38][4], researchers ask cyclists to ride instructed bicycles on road. The bicycles are equipped with camera that records footage to further calculate vehicle velocity, and LiDAR to record distance to the overtaking vehicle. Some works[42][44][12] use data from traffic sensors (e.g. cameras) and analyze the footage from the naturalistic driving setting.

Some experiments[38][30][6] also survey the ego driver and/or the cyclist on their perceived safety immediately after participating in the overtaking during controlled experiments. Afterwards they model relationship between the recorded sensor data and the perceived safety level.

While these experiments provide valuable insights by controlling parameters and investigate into specific factors, they also come with notable limitations.

Firstly, in these controlled overtaking settings, either vehicle driver overtakes a dummy cyclist that moves on a fixed trajectory, with the possible presence of a balloon oncoming vehicle[38]; or the cyclist rides on an instrumented bicycle to record data when they are overtaken by a vehicle. It is morally impossible to conduct experiments that reproduce real vehicle overtaking real cyclist as such high speed interaction is way too dangerous. But this implies potential psychological bias, as the vehicle drivers and the cyclist are aware of the experimental settings, they may unintentionally perform differently than their real-life decisions, resulting in biased results.

Secondly, the test-track experiments took place in a test field, which means highly simplified environment and scenarios. For example, only one oncoming vehicle is used in the experiment, while in real life there might be a few oncoming vehicles that affect the driver's decision. There is a huge gap between the test field and the real driving conditions.

1.2.4 Perceived Safety and Driver Behavior

While many studies have focused on physical indicators of safety such as lateral distance or time-to-collision, a growing number of works also explore how overtaking maneuvers are perceived by cyclists and drivers.

In several controlled experiments [38, 30, 6], participants are asked to rate how safe or unsafe they felt during or after an overtaking event. These subjective ratings are then compared to objective data collected from sensors, such as distance to the cyclist, vehicle speed, and road layout.

This type of research is important because physical safety does not always align with perceived safety. For example, a maneuver with sufficient distance might still feel unsafe if it occurs at high speed or includes sudden steering. Similarly, some close interactions might feel acceptable in low-speed contexts.

Other studies [41, 42] also show that people feel risk differently depending on their experience, habits, or even where they live. Because of this, safety models that only use physical numbers might miss how people actually feel during overtaking.

In this thesis, we aim to incorporate this perspective by designing a scoring system that better reflects perceived discomfort from the cyclist's point of view, using physical measurements but weighting them in a way that matches subjective experience.

1.2.5 Overtaking Phases

The overtaking scenarios are often segmented into different phases for better analysis. Dozza et al.[14], Rasch et al.[39][38] separated the overtaking scenario into four phases: approaching phase, steering away phase, passing phase, and returning phase. Although the naming for the different phases are the same, the specific definition for the phases can be different across different papers, even from the same authors. Generally speaking, approaching phase refers to the period when the ego vehicle driver approaches the cyclist. Steering away phase is the period from when ego driver started steering away to the start of the passing phase. Passing phase is the time when the ego vehicle is passing the cyclist, in some paper, it is 2 meters extended from where the ego vehicle has longitudinal overlap with the cyclist. Rasch et al.[38] extracted different safety metrics in these four overtaking phases that are critical to different types of crash. One major drawback for this four-phase definition is that the start of steering away phase usually requires manual identification, rather than automatically extracting from the signals.

1.2.6 ADAS Support

Some research focus on the safety metrics to support ADAS to avoid potential accidents in overtaking scenarios. Brijs et al.[7] investigated the relationship between lateral distance and time-to-danger with the cyclist to develop a warning feature for ADAS, where time-to-danger refers to the amount of time remaining before the ego vehicle reaches the point where its path starts to overlap longitudinally with the cyclist, potentially leading to a collision. The system identifies the scenario as 'avoidable accident' when lateral distance is lower than 1 meter while time-to-danger is lower than 2 seconds.

This research provides us a safety metric when ego vehicle is approaching the cyclist.

1.2.7 Passing Distance

There are some laws and regulations that restrict the minimum passing distance (MPD) when a motor vehicle is passing (including overtaking) a cyclist. As summarized by Rubie et al.[43]: Queensland in Australia has passed the MPD law in 2016 which requires the MPD to be larger than 1 meters if the ego speed is lower or equal to 60 km/h, and MPD to be no less than 1.5 meters if the ego speed is higher than 60 km/h. Belgium has passed the law that requires MPD to be 1 meter in 1975, and Ontario in Canada has passed the same MPD limit in 2017. France requires the MPD to be 1 meter in urban area, and 1.5 meters else where after 2003. Spain has a MPD limit of 1.5 meters after 2003. 26 states in the USA such as Florida, California, Louisiana, Maryland, Michigan, Minnesota, and Wisconsin have a 3-foot law that requires MPD to be larger than 3 feet (0.9144 meters), dating back to 1973 at earliest.

Although there are laws and regulations that requires the minimum lateral distance while passing a cyclist, some even date back to 50 years ago, not all drivers obey these rules. It is reported that only 64% of the drivers in Spain comply with the 1.5 meters MPD rule[27], 88% in Queensland that comply with the 1 meter MPD law when they are driving slower than 60 km/h, and 79% compliance with the 1.5 meters MPD law at speed greater than 60 km/h[18].

While we can adapt these traffic rules as a "bottom-line" for defining the safe vehicle-cyclist overtakes, we also wonder under what circumstances where the driver chooses to break the law.

1.3 Purpose

Dealing with risky overtaking maneuvers by using real-time warnings or system interventions like ADAS can be difficult. Sometimes drivers ignore the warnings, and interventions may even cause more dangerous situations if the driver does not cooperate. Moreover, drivers complain and increase disbelief when errors were made by the ADAS functions[3][33]. Instead, giving feedback to the driver after a risky overtaking can be a safer and more effective way to help improve driving behavior. In this thesis project, in collaboration with Volvo Cars Corporation (VCC), we aim to analyze VCC's quantitative log data to model and evaluate the riskiness of overtaking scenarios when a vehicle overtakes a cyclist. There are two major applications with our model: post-event driver behavior evaluation system and user-based insurance.

The first primary applications of the developed model is its integration into a post-event driver behavior evaluation system. After an overtaking event occurs, this system will analyze the recorded data and provide automated feedback to the driver. The feedback may include suggestions for safer overtaking practices, warnings about close proximity to vulnerable road users, or commendation for safe driving behavior. This system aims to enhance drivers' situational awareness and promote long-term behavioral improvement through data-driven coaching.

The second application is to support user-based insurance systems[2][21], which fine-tunes insurance premiums based on individual driving behavior, enabled by data. By quantifying the risk levels of overtaking maneuvers, especially near cyclists, the model can provide insurers with objective, behavior-linked metrics to assess driver safety profiles. This encourages safer driving by financially incentivizing responsible behavior and discouraging risky maneuvers. The model's output will contribute to a more nuanced and fair assessment of driver performance. Also provide data support to user-based insurance that adjusts the insurance premium according to driver's behavior.

1.4 Goals

The expected outcome for this thesis work is three-fold.

Firstly, we aim to analyze and process naturalistic driving data. For this, we use data from one of VCC's field data collection sets which has been processed offline by a sensor-fusion-based perception system. With the data, we extract our target scenarios, where the ego vehicle overtakes or passes a cyclist. Passing events are also included as they can be categorized as a special case of overtaking. After identifying these events, we extract key features that describe the overtaking behavior, such as lateral distance, ego vehicle speed, and time to collision with oncoming vehicles. These values are stored on VCC server, they can be queried and fetched for later modeling.

Secondly, we aim to build a safety model that uses the extracted data to predict the overall risk level of an overtaking event to provide feedback to the drivers if they should improve their overtaking maneuver in the future. The safety model incorporates multiple safety metrics in different overtaking phases. Then we evaluate the safety model with VCC's field data to verify if the results align with manual visual inspection.

Thirdly, we aim to identify hard cases that the models may give opposite results than reality, and point out the limitations in the data, such as incomplete data, sensor errors, unusual road users, and unclear driving conditions. Moreover, we aim to provide suggestions for future improvements, such as suggestions for model optimization or pointing out possible solutions to the current limitations.

2

Theory

2.1 Defining Overtaking Maneuvers

In this study, the overtaking maneuver is defined according to the phase-based framework proposed by Dozza et al.[14], As Figure 2.1 shows, cyclist overtaking maneuvers can be split into four phases:

1. **approaching phase:** ego vehicle approaches the cyclist, driver decides which overtaking maneuver to perform.
2. **steering away phase:** ego vehicle steers away to achieve a lateral distance to the cyclist.
3. **passing phase:** ego vehicle passes the cyclist while driving in parallel.
4. **returning phase:** ego vehicle has passed the cyclist and steers back to the lane.

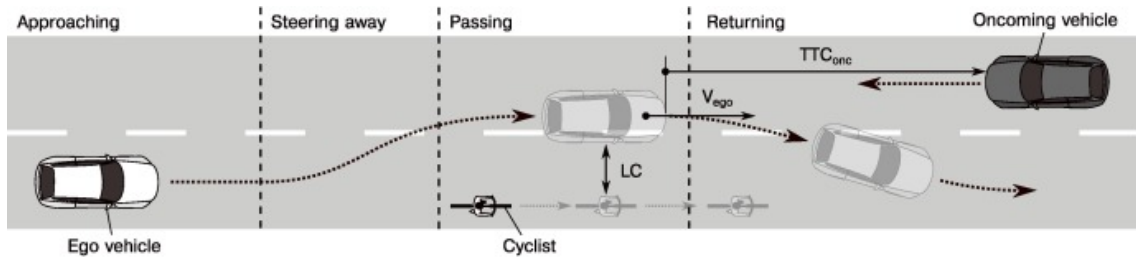


Figure 2.1: Overtaking scenario definition for an exemplary flying maneuver, including overtaking phases and independent variables. LC is the lateral clearance between the ego vehicle and the cyclist, and V_{ego} is the speed of the ego vehicle at the passing moment. TTC_{onc} is the time-to-collision between ego and oncoming vehicles at the passing moment. Source: Rasch et al. [41], *Drivers' and cyclists' safety perceptions in overtaking maneuvers*, 2022.

2.2 Classification of Overtaking Strategies

According to previous research[39], there are mainly two strategies of overtaking cyclists:

- **Flying:** ego vehicle overtakes the cyclist without a clearly deceleration, probably because ego vehicle has a good visibility or complete the maneuver before the oncoming vehicle has reached the cyclist.
- **Accelerative:** ego vehicle brakes and has a clearly deceleration before steering away, probably because ego vehicle has a poor visibility or waiting for the

oncoming vehicle passing first.

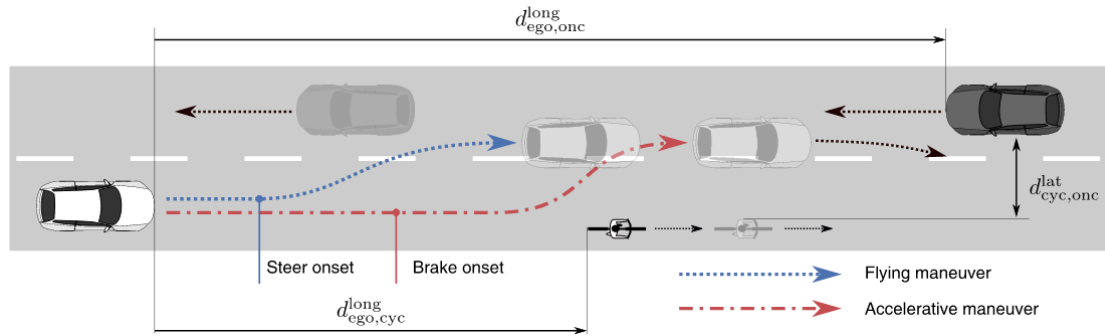


Figure 2.2: Two strategies of overtaking. The blue line indicates a flying maneuver that the ego vehicle (white) steers to overtake the cyclist before oncoming vehicle (black) comes close. The red line indicates a accelerative maneuver that the ego vehicle decelerates to let oncoming vehicle pass before overtaking the cyclist. Source: Rasch and Dozza [39], *Modeling Drivers’ Strategy When Overtaking Cyclists in the Presence of Oncoming Traffic*.

Note that for flying maneuvers, the steering action can be seen as the start of the flying overtaking maneuver, and for accelerative maneuvers, the deceleration (brake) action can be such indication of start.

2.3 Safety Metrics Introduction

To achieve refined modeling and risk assessment of overtaking behavior, we introduce the following key models and metrics (we will introduce how these safety metrics are applied in Section2.5):

Table 2.1: Descriptions of safety-related models and metrics

Model / Metrics	Definition / Description	Note
Variable-Adjusted Minimum Passing Distance (VAMPD)	Country-specific traffic rules.	Most simple, should always follow.
Perceived Risk Score (PRS)	Estimate perceived safety score of driver and cyclist during overtaking.	Determining overtaking strategy is a challenge in our naturalistic dataset.
LD-TTD Risk Index(LTRI)	Real-time warning model inspired by ADAS warning system research	Using lateral distance and time-to-danger as activation criteria.
Minimum Distance Returning (MDR)	Evaluates safety of lane return after overtaking.	Specific used in returning stage

2.3.1 Variable-Adjusted Minimum Passing Distance (VAMPD)

The Variable-Adjusted Minimum Passing Distance (VAMPD) safety metric is based on country-specific traffic rules, requiring the ego vehicle to maintain a minimum lateral distance when overtaking a cyclist. In some places, the required distance also depends on the vehicle’s speed and whether the road is urban or rural. VAMPD takes lateral distance, ego vehicle speed and road type into account, which gives a more realistic and rule-compliant measure of safe passing behavior. For this reason, we refer to this safety metric as Variable-Adjusted MPD instead of simply MPD, as it adjusts the minimum passing distance based on both speed and road type. For the details of the rules in different countries, please see B.1

2.3.1.1 Advantages and Limitations

2.3.1.1.1 Advantages

- **Simple and Intuitive:** Easy to understand and implement, based on clear legal thresholds (e.g., 1.5 meters lateral distance).
- **High Interpretability:** The decision logic is transparent, making it suitable for policy evaluation and public communication.
- **Regulation-Aligned:** Directly reflects compliance with traffic laws, useful for legal or rule-based assessments.
- **Low Data Requirements:** Requires only basic inputs such as lateral distance and vehicle speed.

2.3.1.1.2 Limitations

- **Lacks Contextual Awareness:** Does not account for dynamic traffic factors like oncoming vehicles, road width, or cyclist behavior.
- **Binary Judgment Only:** Can only determine whether a maneuver is legal or illegal, without indicating how risky it is.
- **Limited Flexibility:** Hard to adapt to different countries or regions with varying traffic rules.

2.3.2 Perceived Risk Score

The Perceived Risk Score (PRS) has been calculated using a Bayesian cumulative model developed by Rasch et al. [41], trained on cyclist and driver risk ratings. The previous work set several overtaking cyclist experiments with participants and got a number of function parameters. Therefore, we use the values of model parameters directly in our implementation. The model uses the following logit formulation:

$$a_k - \beta_{LD} \cdot LD_i - \beta_V \cdot V_{ego,i} - \beta_{St} \cdot St_i - \beta_{OP} \cdot OP_i - \beta_{TTC} \cdot TTC_{onc,i} \cdot St_i \cdot OP_i \quad (2.1)$$

We used the extracted scenario features (lateral distance, ego speed, overtaking strategy, presence of oncoming vehicle, and TTC) as input to this model for each overtaking case. The model was applied separately for driver and cyclist perception, yielding two safety scores per scenario.

This model allows us to estimate perceived safety scores from both the driver’s and cyclist’s point of view, which we later use to evaluate and classify the safety level of each overtaking maneuver.

2.3.2.1 Advantages and Limitations

2.3.2.1.1 Advantages

- **Human-centered:** The model is designed to reflect the perceived safety of both drivers and cyclists, making it more aligned with real-world subjective experiences.
- **Quantitative Risk Assessment:** The model provides a continuous risk score beyond binary safe/unsafe judgments.
- **Dual perspective:** The model provides separate safety scores for both the driver and the cyclist, enabling a more comprehensive evaluation of overtaking behavior.

2.3.2.1.2 Limitations

- **Limited generalizability:** The original model was trained on controlled experimental data, which may not fully capture the variability in naturalistic driving environments.
- **Simplified interaction:** The model assumes a single oncoming vehicle and does not account for complex traffic interactions such as multiple oncoming vehicles or curved roads.

- **Static parameters:** Using fixed parameters may reduce sensitivity to contextual variations present in real-world data.
- **Strict data compatibility:** Requires structured and clean data inputs, which may limit its use in diverse or real-time environments.

2.3.3 LD-TTD Risk Index(LTRI)

The LD-TTD Risk Index(LTRI) safety metric in this study is inspired by the work of Brijs et al. [7], which investigated the effects of an advanced driver-assistance system on the safety of cyclist overtaking maneuvers. Their research demonstrated that timely warnings based on lateral distance and time-to-danger metrics can significantly improve driver behavior and reduce risky overtaking events.

The *Time-to-Danger (TTD)* is a surrogate safety measure adapted from the classical Time-to-Collision (TTC), but extended to apply to non-collision overtaking scenarios[7]. While TTC is defined as the time until impact if two road users remain on a collision path[19], TTD is used to estimate how soon a vehicle will become laterally aligned with a cyclist—an important moment in assessing the severity of the maneuver.

Mathematically, TTD is defined as:

$$TTD = \frac{Dist_{vh-c}}{Speed_{vh} - Speed_c} \quad (2.2)$$

Where:

- $Dist_{vh-c}$ is the longitudinal distance (in meters) between the vehicle’s front bumper and the rear wheel of the cyclist.
- $Speed_{vh}$ is the longitudinal speed (in m/s) of the vehicle.
- $Speed_c$ is the longitudinal speed (in m/s) of the cyclist.

TTD is only defined when $Speed_{vh} > Speed_c$, i.e., when the vehicle is approaching the cyclist from behind. [19]

The warning strategy operates under the following logic:

Table 2.3: Risk classification logic of the LD-TTD Risk Index (LTRI)[19]

Lateral Distance	4.5 s > TTD ≥ 3 s	3 s > TTD ≥ 2 s	TTD < 2 s
LD ≥ 1.5 m	Normal driving	Normal driving	Normal driving
1.5 m > LD ≥ 1.0 m	Normal driving	Danger	Danger
LD < 1.0 m	Normal driving	Danger	Avoidable Accident

2.3.3.1 Advantages and Limitations

2.3.3.1.1 Advantages

- **Balanced complexity:** Simple enough to compute in real time, but more informative than one-dimensional safety metrics.
- **Phase-specific suitability:** It works especially well in the approaching phase, where early risk detection is critical.
- **ADAS-inspired:** The design of LTRI is inspired by ADAS warning systems, making it naturally compatible with future applications in real-time driver assistance.

2.3.3.1.2 Limitations

- **Lacks advanced complexity:** While its simplicity is an advantage, LTRI does not capture more complex risk patterns that may involve multi-agent interactions.

2.3.4 Minimum Distance Returning (MDR)

The MDR safety metric is inspired by the work of Rasch et al. [38], which analyzed how oncoming traffic and cyclist lane position influence driver behavior during overtaking. One of their key findings was that the timing and trajectory of the vehicle's return to its original lane (i.e., the steering back phase) can significantly affect cyclist safety.

2.3.4.1 Advantages and Limitations

2.3.4.1.1 Advantages

- **Phase-specific suitability:** Focuses on the steering back phase, which is often overlooked but important for evaluating the full overtaking maneuver.
- **Simplicity:** The metric is easy to compute once the overtaking phases are identified.

2.3.4.1.2 Limitations

- **Lacks complexity:** Does not capture more complex risk patterns that may involve multi-agent interactions.

2.4 Scenarios Classification and Scope

We classify the passing scenarios on the road into three categories: **Overtaking**, **Drive-by** and **Meeting**.

Since meeting scenarios are quite different from our topic of drive-by and overtaking (i.e. the ego vehicle approaches cyclists from behind), as the oncoming cyclist is much more aware of the situation and is often more proactive in decision-making during meeting [5], we decided to exclude them from the analysis in the scope of our thesis work.

However, we have kept this type of scenario in the scenario extractor output to provide potentially useful data for future research and to support other researchers or engineers at VCC who may be interested in investigating these scenarios further.

2.5 Phase Definitions and Model Applications

To more accurately assess the risks during overtaking, we divide passing scenarios into two types: **Drive-by** and **Overtaking**, and define their respective phases and each applying different related safety metrics.

2.5.1 Drive-by Scenario Modeling

Drive-by refers to the ego vehicle passing the cyclist without obvious steering maneuvers. This scenario often occurs when the cyclist is in an independent cycling lane, and the ego vehicle stays in its lane, making oncoming vehicles usually irrelevant.

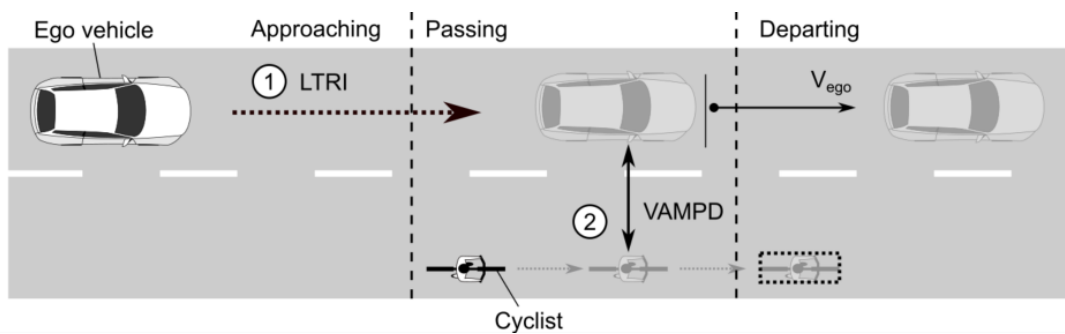


Figure 2.3: Passing scenario and related safety metrics:(1)LD-TTD Risk Index (LTRI) applied in approaching phase,(2)Variable-Adjusted Minimum Passing Distance (VAMPD) applied in passing phase

We define three phases for passing scenario:

2.5.1.1 Approaching Phase

- Time: 5 seconds before passing
- Focus: Ego speed and lateral distance with the cyclist
- Purpose: Avoid front-rear crash, considering driver's reaction time
- Safety Metrics Application: LD-TTD Risk Index(LTRI)

2.5.1.2 Passing Phase

- Time: The period during which ego vehicle has longitudinal overlap with the cyclist.
- Focus: Ego speed and lateral distance with the cyclist
- Purpose: Avoid side crash
- Safety Metrics Application: VAMPD

2.5.1.3 Departing Phase

- Time: A few seconds after passing
- Since oncoming vehicles are irrelevant, no focus in this phase

2.5.2 Overtaking Scenario Modeling

Overtaking refers to the ego vehicle exhibiting steering maneuvers and overtaking the cyclist while possible oncoming vehicles are present. This scenario has higher dynamic interaction complexity and potential risk.

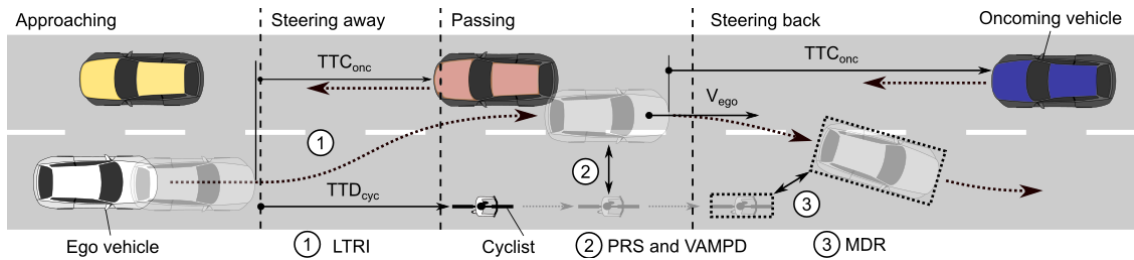


Figure 2.4: Phases and metrics in overtaking scenario with multiple oncoming vehicles present. Since multiple oncoming vehicles may appear throughout different phases of the overtaking scenario (e.g., yellow, red, and blue vehicles), the Time to Collision (TTC) with the respective oncoming vehicle should be evaluated continuously as each vehicle comes into view. Overtaking scenario and related safety metrics: (1)LD-TTD Risk Index (LTRI) applied in approaching phase and steering away phase,(2)Perceived Risk Score (PRS) and Variable-Adjusted Minimum Passing Distance (VAMPD) applied in passing phase,(3)Minimum Distance Returning (MDR) applied in steering back phase

We define four phases for overtaking scenario similar to Marco and Alex’s perceived safety paper[41]:

2.5.2.1 Approaching Phase

- Time: 5 seconds before passing
- Focus: Driver’s decision of when to initiate steering, i.e., Time to collision to the oncoming vehicles, possibly waiting for some oncoming vehicles to pass first
- Safety Metrics Application: LD-TTD Risk Index(LTRI)

2.5.2.2 Steering Away Phase

- Time: From the steering start to passing start
- Focus: Time to danger, ego speed, and lateral distance with the cyclist
- Safety Metrics Application: LD-TTD Risk Index(LTRI)

2.5.2.3 Passing Phase

- Time: The period during which ego vehicle has longitudinal overlap with the cyclist.
- Note: With possible oncoming vehicles are present, possibly also having longitudinal overlap with a passing oncoming vehicle (squeeze through maneuver)
- Focus:
 - Ego speed and lateral distance with the cyclist
 - TTC with oncoming vehicles
 - Ego speed and lateral distance with the passing oncoming vehicle
- Safety Metrics Application: Perceived Risk Score (PRS), Variable-Adjusted Minimum Passing Distance (VAMPD)

2.5.2.4 Steering Back Phase

- Time: 5 seconds after passing
- Focus: TTC with oncoming vehicles, Euclidean distance to the cyclist
- Safety Metrics Application: Minimum Distance Returning (MDR)

3

Methods

We first extract the scenarios where the ego vehicle overtakes cyclists, whatever the overtaking way it takes, from VCC’s sensor-fused logged data using the *Scenario Extractor* tool. These include both overtaking and drive-by scenarios.

After identifying relevant scenarios, we perform phase extraction to segment each scenario into meaningful temporal phase. Such segmentation provides deeper insight into when and how safety-critical situations occur, which allows a more comprehensive assessment of overtaking safety.

Finally, we compute and output various contextual and dynamic signals during each phase, which serve as the input for our safety models and metrics. Unlike previous methods that focus only single phase, our phase-based safety analysis offers a broader and more accurate view of overtaking safety.

3.1 Data Description

This study is based on VCC’s log data collected from instrumented test vehicles driven under real-world conditions in different cities around the globe.

The test vehicles that are equipped with multiple sensors collect information while driving in real-world traffic, then the logs are stored and uploaded to VCC server, where data pipelines process these raw outputs from the different sensors, using technologies like sensor fusion to produce a timestamp-aligned, preprocessed and structured representation of the driving history. We refer to this single driving log as a *stream*, and the preprocessed log data as *sensor-fused data* (SF data). We can access various information (commonly referred to as *signals* in the field), on the camera footage, traffic objects (including ego vehicle and other various types of road users), lane markings, road information, traffic signs and road condition[1].

VCC also performs simulation (resim) on the SF data and virtually visualizes the scenarios in a bird-eye view and displays various signals in a user-friendly visualization system. Users can easily visually inspect a certain scenario in this system. The data we had access to has thousands of streams where each stream lasts between 30 minutes and 3 hours.

3.2 Scenario Extractor

We used and contributed to an internal tool called *Scenarios Extractor* at VCC. This tool processes sensor-fused log data and identifies specific driving scenarios based on predefined and carefully designed conditions. It also allows the extraction of

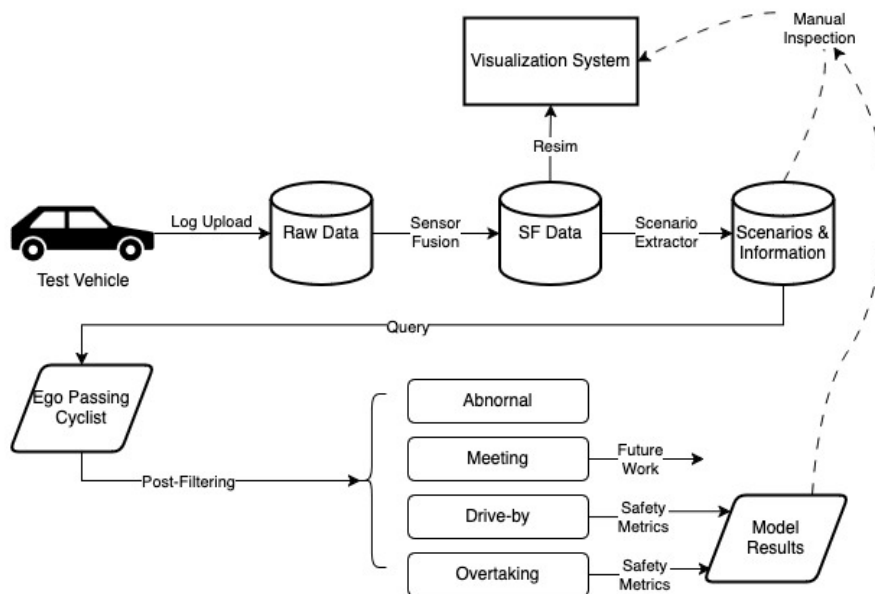


Figure 3.1: Data processing flow in the scope of this thesis work.

custom, derived values from the available signals, enabling further analysis in these driving scenarios.

3.2.1 Finder

In search for the scenarios in which the ego vehicle passes a cyclist. We first tried to find some useful results from the existing output of Scenario Extractor. However, although we have tried many different combinations of the query conditions, these findings are not accurate enough for us. So we implemented a new finder in the Scenario Extractor tool to identify our specific ego passing cyclist scenario, namely *EgoPassingCyclistStateFinder*.

We used the following criteria to identify an ego passing cyclist scenario:

- The overtaken object is a bicycle or unknown.
- The center-to-center lateral distance with the overtaken object is less than 5.25 meters (which means the side-to-side lateral distance is around 3.9 meters).
- Ego vehicle has longitudinal overlap with the object.
- Ego vehicle should be moving forward and move faster than the overtaken object.
- The object should be heading relatively parallel to the ego vehicle (absolute relative heading angle less than 60 degrees).

These criteria were designed to capture the overtaking scenarios, with a focus on balancing between completeness and relevance to our research topic. Below we explain each criterion in detail and provide reasoning for the selected thresholds.

In the first condition, we specify the object class to be bicycle or unknown. Including the unknown class is a common practice when using this tool, in order to also include the periods where the sensors are unable to accurately identify the type of the object, thus including as many relevant scenarios as possible.

In the second condition, we are filtering out the cyclists that are too far away from us, as they are usually not in the scope of the overtaking. We deliberately set the final center-to-center lateral distance threshold to 5.25 meters, which is 1.5 times the average lane width. If we set this threshold too small, it will result in reduced number of findings, while setting this threshold too large will result in more irrelevant cases. It is meant to include a wide range of possible overtaking maneuvers, from close passes to more cautious ones, as validated during manual inspection on the passing scenarios in an earlier stage of implementing this tool. Given that the test vehicle is 1.923 meters in width and bicycles are considered 0.75 meters in width, this corresponds to a side-to-side distance of approximately 3.9 meters.

In the third condition, we require the vehicle-cyclist longitudinal overlap to ensure that the ego vehicle is truly in an overtaking situation rather than driving ahead or following the cyclist. We use this criteria to identify the start and the end timestamp for our passing phase.

Similarly in the fourth condition, the constraint that the ego vehicle must be moving forward and faster than the object ensures the proper directionality of the interaction. This helps exclude irrelevant events such as stationary or reversing maneuvers, or situations where the cyclist overtakes the vehicle that we often find in the preliminary versions of our scenario extractor.

Finally, bounding the relative heading between the ego vehicle and the cyclist to within 60 degrees ensures that both are generally moving in the same direction. This is a relatively broad constraint that allows for slight road curvature or imperfect sensor estimation; if stricter heading directions are desired, such as retaining only near-parallel ones, additional filtering can be applied in the next steps.

Overall, these are carefully designed criteria that balance the number of findings and the relevance of the findings in our custom *EgoPassingCyclistStateFinder*. This finder finds the period of time in a stream where the ego vehicle passes a cyclist.

3.2.2 Phases Extraction

The finder in Scenario Extractor returns the time period from passing start and passing end. We defined a pre-passing phase and a post-passing phase, which are defined as 5 seconds before and after passing respectively.

Then we look into the steering maneuver in the pre-passing phase, and define the distinction between approaching phase and steering away phase as 0.5 seconds before the steering wheel angle reaches the maximum value. Thus, the entire scenario is divided into four phases, namely approaching, steering away, passing and post-passing phases (departing or steering back). These four phases defined in scenario extractor align with the phase definitions in the drive-by and overtaking scenarios.

3.2.3 Output

We then define, calculate and output the various information relevant to our research within the periods. For example, we calculate the average ego longitudinal velocity and the minimum lateral distance during the passing, along with the country/state information processed from GPS coordinates and the road type; together, these

outputs enable the calculation of our VAMPD model originated from the country-specific traffic rules.

The full information of the output signals, including names, types, units, usage and notes can be found in the tables in Appendix C.

3.2.4 Small Batch Run

We began by running the scenario extractor on a small-scale dataset consisting of 200 individual streams. The streams were randomly selected from a manually extracted list of streams where there are bicycles present on tertiary road, to increase the probability of finding simpler and more suitable scenarios.

The key idea for this small batch run was to allow us to have the possibility for detailed manual inspection, enabling us to validate the correctness of the scenario extractor's output, as well as to evaluate the models and the metrics.

3.2.5 Full-Scale Run

After the small batch run, we ran the scenario extractor on all available 8300 streams. This output enabled us to conduct thorough data analysis and apply our models and metrics on it.

3.3 Post Filtering and Scenario Classification

We perform post-filtering process to the output from scenario extractor. The aim for this process is to filter out abnormal data, such as ego vehicle moving too slowly or cyclist has abnormal heading, and further classify the ego passing cyclist scenarios into the following types:

- **Meeting:** Ego vehicle meets the cyclist that comes from the opposite direction.
- **Drive-by:** Ego vehicle drives by the cyclist without steering away maneuver intended to avoid the cyclist or to create a larger lateral distance. Refer to the figure 3.2 below.
- **Overtaking:** Ego vehicle overtakes the cyclist with steering away maneuver intended to avoid the cyclist or to create a larger lateral distance. Refer to the figure 3.3 below.

For the meeting case, we filtered cases with negative object velocity, which indicates that the cyclist is coming towards ego vehicle. For the overtaking case, we use combined conditions, i.e. road should be straight (average road curvature less than 0.57 degrees), there should be lateral distance change between steering start and passing start (greater than 0.5 meters). And we categorize the rest cases into drive-by scenarios.

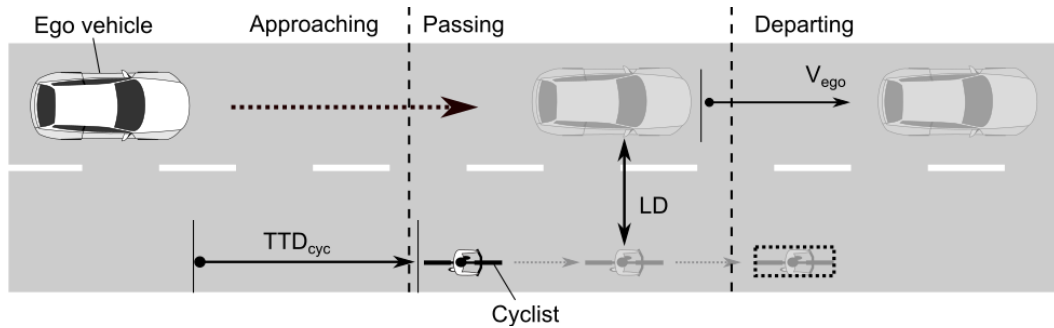


Figure 3.2: The figure shows the definition of drive-by scenarios, including phases and independent variables. LD is the lateral distance between the ego vehicle and the cyclist, and V_{ego} is the speed of the ego vehicle. TTD_{cyc} is the time-to-danger between ego vehicle and cyclist before passing.

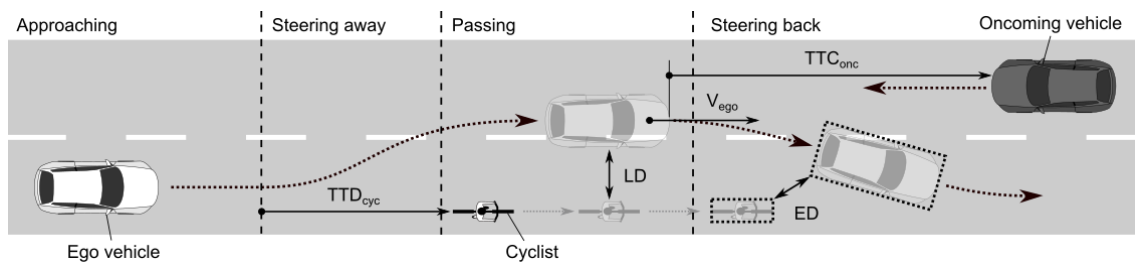


Figure 3.3: The figure shows the definition of overtaking scenarios, including phases and independent variables. LD is the lateral distance between the ego vehicle and the cyclist, and V_{ego} is the speed of the ego vehicle. TTC_{onc} is the time-to-collision between ego and oncoming vehicles at the passing moment. TTD_{cyc} is the time-to-danger between ego vehicle and cyclist before passing. ED is the Euclidean distance between ego vehicle and cyclist at the steering back phase.

3.4 Application of Safety Metrics

After classifying the cases, we chose to investigate into the drive-by cases and the overtaking cases, and leave meeting cases for future work. We applied different safety metrics on the two cases.

3.4.1 Variable-Adjusted Minimum Passing Distance (VAMPD) Implementation

To evaluate overtaking behavior using the VAMPD metric, we compare the lateral distance between the ego vehicle and the cyclist against country-specific legal thresholds. These thresholds depend on factors such as vehicle speed and road type, and are summarized in Appendix B.1.

3.4.1.1 Computation

For each overtaking scenario, we extract the minimum lateral distance during the passing maneuver. This distance is then evaluated against the legal passing distance, determined as follows:

- If countries or regions have their own MPD rules, (e.g., Australia, Ireland, Germany, UK), the rule is directly applied and we will dynamically adjust the threshold based on the ego vehicle’s speed and available road metadata.
- For countries or regions without specific MPD rules, we assume a default threshold of 1 meter for safety, consistent with international best practices for protecting vulnerable road users.

3.4.1.2 Binary Safety Evaluation Criterion

We apply a binary classification:

- **Safe:** If the vehicle’s passing distance follows the rules of the country, or meets our assumed safety distance when no law exists.
- **Unsafe:** If the vehicle does not meet the legal or assumed safety distance.

This approach matches the idea of VAMPD as a rule-based metric: any violation of legal or safety limits means the maneuver is unsafe.

3.4.2 Perceived Risk Score Implementation

We implemented the Bayesian cumulative model using the extracted scenario-level features: lateral distance, ego vehicle speed, overtaking strategy, presence of oncoming vehicle, and time-to-collision (TTC). These inputs were derived from each overtaking case (see Table 3.2).

Variable	Unit	Values/Options	Notes/Annotations
Lateral distance	m	–	–
Ego vehicle speed	km/h	–	Average ego longitudinal velocity
Overtaking strategy	–	Accelerative (OP = 0) / Flying (St = 1)	0: Ego vehicle slowed before overtaking. 1: Overtake without slowing.
Presence of oncoming vehicle	–	Absent (OP = 0) / Present (OP = 1)	–
TTC to oncoming vehicle(s)	s	–	For flying maneuvers with oncoming vehicle present

Table 3.2: Variables and annotations for overtaking scenarios

3.4.2.1 Model Parameterization

We did not retrain the model but instead used the fixed population-level median values of each parameter from the original study (see Table 3.3).

Table 3.3: Median values of model parameters for driver and cyclist perceived safety models, adapted from [41]. These values were used directly in our implementation.

Parameter	Unit	Driver Model (Median)	Cyclist Model (Median)
a_1	NA	-0.59	-3.55
a_2	NA	1.58	-0.19
a_3	NA	3.25	1.24
a_4	NA	4.56	1.61
a_5	NA	6.13	NA
a_6	NA	9.55	NA
β_{LD}	1/m	-0.26	-4.00
β_V	s/m	-0.07	0.05
β_{St}	NA	1.72	0.64
β_{OP}	NA	3.32	5.37
β_{TTC}	1/s	-0.31	-0.13

In the original model, a group-level effect u_{ID} is included to account for differences

between participants:

$$u_{\text{ID}} \sim \mathcal{N}(0, \sigma_{\text{ID}}^2)$$

However, in our implementation, we do not include this participant ID term, as our dataset does not contain individual driver or cyclist identifiers. Therefore, our version of the model assumes fixed effects across all overtaking events.

3.4.2.2 Scoring Method Modification

We made improvements in the statistical analysis. In the original study, the final perceived safety score was determined by selecting the score level with the highest predicted probability. However, this approach may overlook the distributional nature of the model output and overemphasize a single outcome.

To improve robustness and better reflect the full risk profile, we modified the aggregation method by computing a **weighted average score**, where each score level is weighted by its predicted probability. This approach captures the expected value of the perceived risk and provides a smoother, more representative measure of safety perception.

Formally, the final score is computed as:

$$\text{Weighted Average Score} = \sum_{k=1}^K k \cdot P(\text{score} = k)$$

Then scale the scores back to the original range by applying a min-max scaler to all weighted average scores:

$$X' = a + \frac{(X - X_{\min})(b - a)}{X_{\max} - X_{\min}}$$

Where X denotes the original value of the feature before scaling. X_{\min} denotes the minimum value of the feature in the dataset. X_{\max} denotes the maximum value of the feature in the dataset. X' denotes the normalized value after applying Min-Max scaling, which will lie in the range $[a, b]$. For the driver’s model, $a=1$ and $b=7$. For the cyclist’s model, $a=1$ and $b=5$.

This enhancement allows for more nuanced comparisons across scenarios and avoids abrupt changes caused by small shifts in probability mass.

3.4.2.3 Safety Evaluation Criterion

For each scenario, we calculate a predicted weighted average score from both the driver’s and cyclist’s perspectives. The driver’s risk score ranges from 1 (very low risk) to 7 (very high risk), while the cyclist’s risk score ranges from 1 (very low risk) to 5 (very high risk). According to the results, these scores reflect the safety level of each scenario.

3.4.3 LD-TTD Risk Index(LTRI) Implementation

The LD-TTD Risk Index (LTRI) is calculated based on the formula introduced in the theory section 2.3.3. For each overtaking scenario, we extract the required

variables to compute the Time-to-Danger (TTD), then combine it with the lateral distance (LD).

3.4.3.1 Safety Evaluation Criterion

Risk levels were classified based on the thresholds in Table 2.3. Events were then labeled as normal driving, danger, or avoidable accident accordingly.

3.4.4 Minimum Distance Returning (MDR)

Minimum Distance Returning is defined as the closest Euclidean distance with the cyclist during the returning phase.

3.4.4.1 Safety Evaluation Criterion

We define a Minimum Distance Returning criterion that in any case, the Euclidean distance to the cyclist should always be larger than 1 meter during the returning phase. This is to avoid the rear-front crash with the cyclist, and avoid close interactions that may raise the perceived risk level for the cyclist.

4

Results

4.1 Results on a Small Batch

4.1.1 Data Description

In the small batch of 200 streams, the scenario extractor returned information (Refer to Appendix C for detailed information on the output names and their usage) about 1184 scenarios where ego vehicle passed a cyclist.

We then applied post filtering process (refer to Section 2.5), removing 53 abnormal cases, resulting in 1131 valid scenarios.

Next, we identified the types for the passing. We categorized the results into three categories as defined in the previous section 3.3:

As a result, among the remaining 1131 scenarios:

- **615 (54.4%)** were classified as *meeting*,
- **472 (41.7%)** were classified as *drive-by*,
- **44 (3.9%)** were classified as *overtaking*.

4.1.2 Comparison on Perceived Risk Score Model Before and After Scoring Method Modification

We first applied the Cyclist’s Perceived Risk Score model on the 44 overtake data during the passing phase, and compared the results between using the original model and the altered weighted model.

Using the original arg-max scoring method, the cyclist scores have an average score of 1.54, with standard deviation of 0.89. As for the weighted scoring method, the average is 1.92 and the standard deviation is 1.03.

As the histograms in Fig 4.1 shows, the original scoring method rarely gives prediction of score 3 and score 4. On the contrary, our weighted scoring method has less results on the low and high scores and has much more results on the mid-range scores.

4. Results

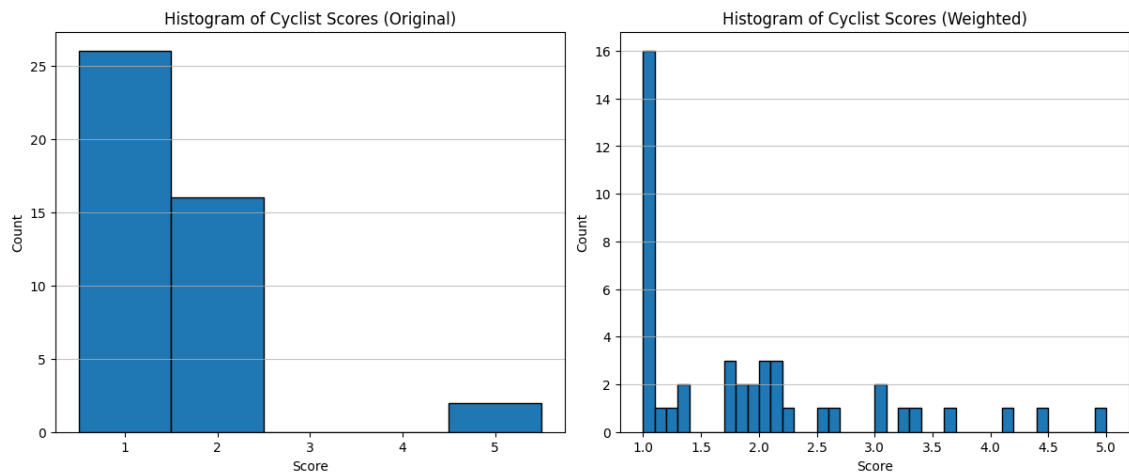


Figure 4.1: Comparison of predicted cyclist perceived risk scores from the original scoring method and the altered weighted scoring methods on a 44-sample batch. The altered weighted model shows fewer low and high scores, with more mid-range predictions.

As Fig 4.2 shows, the distribution for the scores are more smooth. For example, the data point at 28 km/h in ego velocity and 0.78 meters in lateral distance. It has a score of 5 in the original method while all its near data points have a predicted score of 2, indicating a sudden change. On the contrary, this data point has a predicted score of 5.0 in the weighted scoring method, with its surrounding data points has score ranging from 2.09 to 3.38, resulting in smoother transition and less sudden change. To summarize, the weighted model produces smoother score transitions between neighboring data points, highlighting its improved consistency over the original approach.

One thing worth mentioning is the data point that has the weighted score of 4.18 (Ego velocity 37 and lateral distance 1.2), whose adjacent points has much lower scores. We manually inspected this data point. In this case, the ego vehicle showed a flying maneuver, with oncoming vehicle present at a time-to-collision of 12.3 seconds. Comparing to its neighboring points, it is the only point that has oncoming vehicle present. It is reasonable that this point has a higher score using the weighted scoring method, as the presence of the oncoming vehicle increases the risk.

Similarly for the data point at the bottom, that has almost 0 meters in lateral distance. It is scored 5 by the original scoring method, while it a score of 4.49 in the weighted method. It is because that the ego velocity is low, the weighted score resulted in less significant risk level in the model. This would be a drawback for using the weighted scoring method, that the risk level can be downplayed in some cases due to the scaling.

We also investigated the difference between the weighted and original scoring methods. As Fig 4.3 shows, the weighted method tends to output much higher scores than the original model, while having 29 cases (66%) that has nearly same predictions with score difference lower than 0.5.

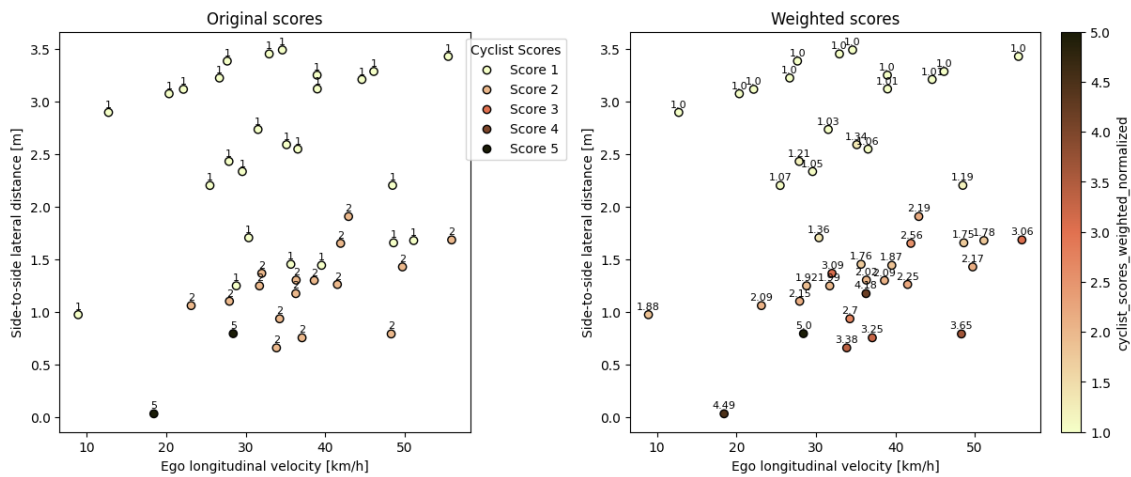


Figure 4.2: Comparison of original and weighted cyclist scoring on small batch data w.r.t ego longitudinal velocity and side-to-side lateral distance.

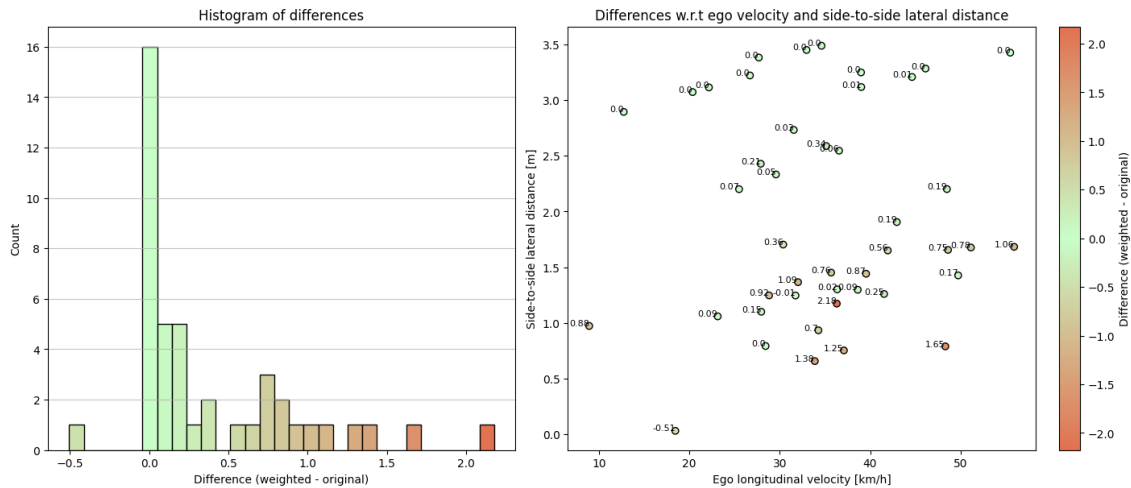


Figure 4.3: Comparison of score differences between the weighted and original models on a small batch dataset. The histogram (left) shows the distribution of score differences, while the scatter plot (right) relates these differences to ego velocity and lateral distance.

As for the Driver’s Perceived Risk model, we are only getting prediction scores 1 and 2 on this small-batch data, so we omit the comparison on the original and the weighted method here.

4.1.3 Drive-by Scenarios

4.1.3.1 LTRI

Among 472 drive-by scenarios, 339 (71.8%) scenarios are identified as normal driving, 80 (17.0%) scenarios are identified as danger, and 53 (11.2%) scenarios are identified as avoidable accident.

4.1.3.2 VAMPD

Among 472 drive-by scenarios, 31 (6.6%) scenarios take place in the countries or cities that has a MPD rule, in which 9 (29%) scenarios are breaking the MPD rule. Considering the default minimum lateral distance to be 1 meters for the areas that do not yet have a relevant traffic rule, among the 441 (93.4%) scenarios, 382 (86.8%) are considered safe, while 59 (13.2%) of them are considered unsafe.

4.1.4 Overtaking Scenarios

We applied the four safety metrics on the overtaking scenarios and did manual inspection on all of the 44 scenarios and rated the overall risk score (considering both drive and the cyclist, ranging from 1 to 5) in the different phases.

4.1.4.1 LTRI

Among the 44 overtaking scenarios, 22 (50%) scenarios are identified as normal driving, 11 (25%) scenarios are identified as danger and 11 (25%) scenarios are identified as avoidable accident.

The results highly align with our manual inspection. During manual inspection, higher risk scores were given when the ego vehicle steered late and has closer lateral distance with the cyclist before passing. All the scenarios that are identified as avoidable accident were manually rated a high risk score (4 or 5).

4.1.4.2 VAMPD

Among the 44 overtaking scenarios, only 1 scenario takes place in the countries or cities that has a MPD rule, which is breaking the MPD rule. Considering the default minimum lateral distance to be 1 meters for the areas that do not yet have a relevant traffic rule, among the rest 43 scenarios, 36 (83.7%) are considered safe, while 7 (16.3%) of them are considered unsafe.

4.1.4.3 PRS

We first summarize the relevant parameter distributions in the PRS model.

Table 4.1: Summary of variables in the overtaking dataset ($N = 44$). All continuous variables are shown as mean (SD); all categorical variables as number of samples per level (percentage).

Variable (unit)	Overtaking Data, $N = 44$
Lateral distance (m)	1.97 (0.96)
Ego velocity (km/h)	34.95 (10.54)
Overtaking strategy	
Accelerative	2 (4.5%)
Flying	42 (95.5%)
Presence of oncoming vehicle	
Absent	36 (81.8%)
Present	8 (18.2%)
TTC to oncoming vehicle (s)	9.85 (2.21)
Valid counts (flying maneuvers with oncoming vehicles present)	8 (18.2%)

As we discussed in the previous subsection, the score from the driver model only has score 1 and score 2, meaning that none of the overtaking scenarios are risky from the driver’s perspective. So we use the original scoring system for driver model, and use the weighted scoring for the cyclist model. The result is shown in the histogram Fig 4.4.

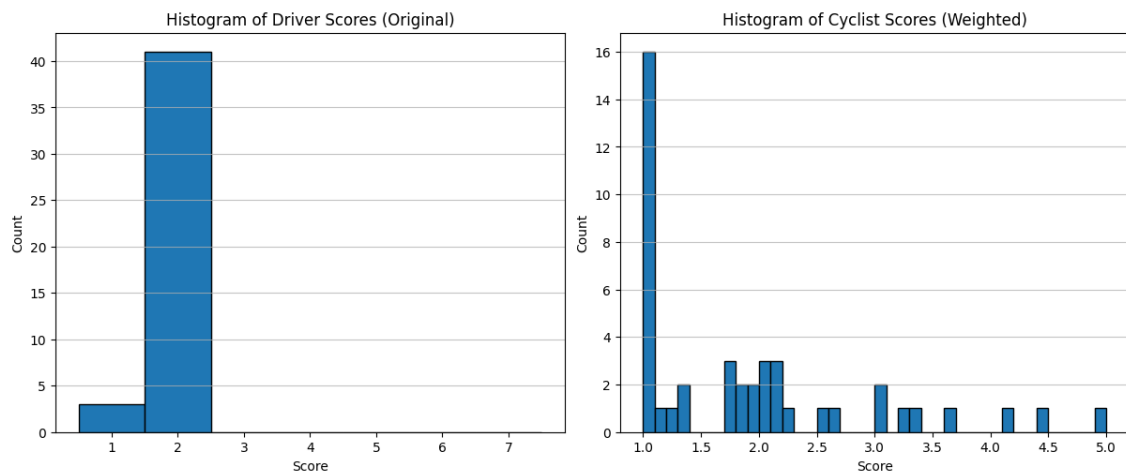


Figure 4.4: Score Distribution for the perceived risk score metric on the small batch dataset. Using original scoring method for the driver model and weighted scoring method for the cyclist model.

4.1.4.4 MDR

Among the 44 overtaking scenarios, the Euclidean distance with the cyclist has an average of 2.52 meters with a standard deviation of 0.98.

We chose 1 meter as the safety threshold, which is about 1.5 standard deviations below the average, meaning that around 7% of the cases would be considered unsafe assuming the data follows the Gaussian distribution. This makes it a simple but strict rule to catch cases where the vehicle gets too close to the cyclist during the return phase.

Fig 4.5 shows the distribution of the MDR.

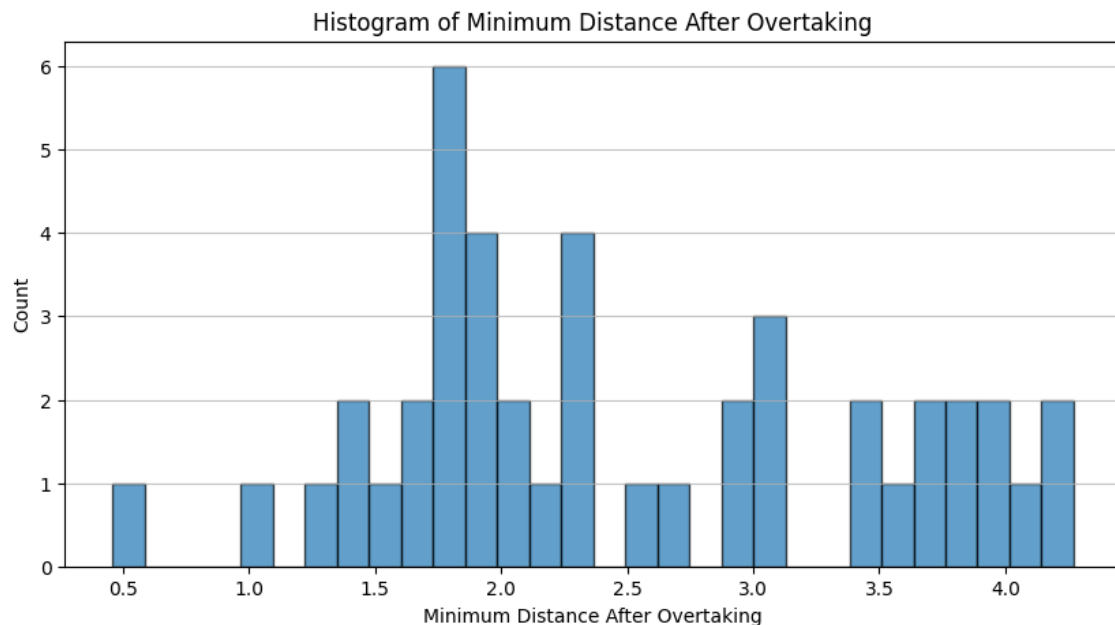


Figure 4.5: Distribution of Minimum Distance Returning (MDR) values calculated across 44 overtaking maneuvers, indicating how closely the vehicle merges back after passing the cyclist.

With the threshold of 1 meter, only one overtaking scenario is considered unsafe with a MDR at 0.46 meters.

We did manual inspection on this unsafe overtaking scenario, but unfortunately there is no cyclist found in the footage, it appeared to be a sensor error that misidentified a small sight-seeing bus with large open windows to be the cyclist. But anyway, having 0.46 meters in minimum Euclidean distance with a small bus after overtaking is still very unsafe.

4.2 Results on Full Dataset

4.2.1 Data Description

In this full-scale run on thousands of streams, the scenario extractor returned information of scenarios where ego vehicle passes cyclist. We filtered out around 5%

abnormal scenarios, with around 95% valid scenarios remaining, among which:

- **51.9%** were classified as *meeting*,
- **43.6%** were classified as *drive-by*,
- **4.5%** were classified as *overtaking*.

4.2.2 drive-by Scenarios

4.2.2.1 LTRI

Among the drive-by scenarios, 74.7% of the scenarios are identified as normal driving, 14.7% of the scenarios are identified as danger, and 10.7% of the scenarios are identified as avoidable accident.

4.2.2.2 VAMPD

Among the drive-by scenarios, 5.8% of the scenarios took place in the countries or cities that has a MPD rule, in which 31.5% of the scenarios were breaking the MPD rule. Considering the default minimum lateral distance to be 1 meters for the areas that do not yet have a relevant traffic rule, among these 94.2% of the scenarios, 11.4% of them were considered unsafe.

Table 4.3: Summary of VAMPD results on drive-by scenarios in full dataset. All variables are shown as percentage.

VAMPD Result	drive-by Data
In countries with MPD rule	5.8%
Obeyed	68.5%
Broke	31.5%
In countries without MPD rule	94.2%
Obeyed	88.6%
Broke	11.4%

In addition to the country specific rules, we also applied the Australian rule on all the data. The Australian rule can requires minimum lateral distance when passing to be 1 meter if speed is lower than 60 km/h, otherwise it should be 1.5 meters.

As a result, 12.2% of the cases in the drive-by scenarios were identified as unsafe, while the rest 87.8% were identified as safe.

4.2.3 Overtaking Scenarios

4.2.3.1 LTRI

Among the overtaking scenarios, 58.4% of the scenarios were identified as normal driving, 23.8% of the scenarios were identified as danger, and 17.8% of the scenarios were identified as Avoidable Accident.

4.2.3.2 VAMPD

Among the overtaking scenarios, 5.5% of the scenarios took place in the countries or cities that has a MPD rule, in which 34% of the cases are breaking the MPD rule. Considering the default minimum lateral distance to be 1 meters for the areas that do not yet have a relevant traffic rule, among the rest 94.5% of the scenarios, 11.5% of them were considered unsafe.

Table 4.5: Summary of VAMPD results on overtaking scenarios in the full dataset. All variables are shown as percentage.

VAMPD Result	Overtaking Data
In countries with MPD rule	5.5%
Obeyed	65.1%
Broke	34.9%
In countries without MPD rule	94.5%
Obeyed	88.5%
Broke	11.5%

4.2.3.3 PRS

We first summarize the relevant parameter distributions in the PRS model as Table 4.7 shows.

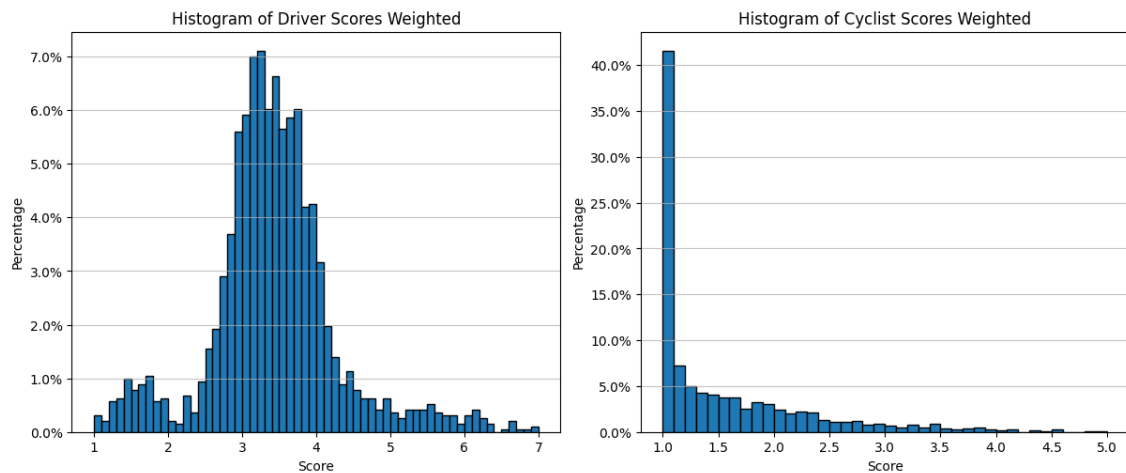


Figure 4.6: Results of Perceived Risk Score distribution on full-scale overtaking dataset.

Table 4.7: Summary of variables in the full-scale overtaking dataset. All continuous variables are shown as mean (SD); all categorical variables as percentage.

Variable (unit)	Overtaking Data
Lateral distance (m)	2.05 (0.88)
Ego velocity (km/h)	34.26 (12.14)
Overtaking strategy	
Accelerative	8.3%
Flying	91.7%
Presence of oncoming vehicle	
Absent	82.3%
Present	17.7%
TTC to oncoming vehicle (s)	9.15 (2.02)
Valid percentage (flying maneuvers with oncoming vehicles present)	15.7%

We use the weighted scoring method for the both models. The result of is shown in the histogram in Fig 4.6.

Since the driver score is more sensitive to ego velocity and TTC with oncoming vehicle, and the cyclist score is more sensitive to ego velocity and lateral distance, we also plot a scatter plot showing the transition of scores, as Fig 4.7 shows.

4. Results

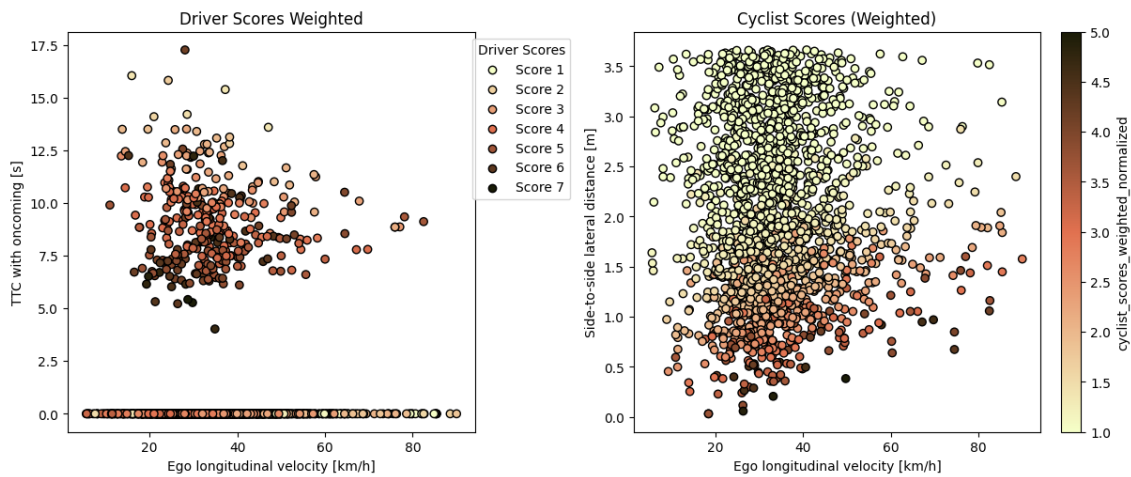


Figure 4.7: Results of Perceived Risk Score in the full-scale overtaking dataset using the weighted method, with respect to ego velocity and TTC with oncoming vehicle (0 if no oncoming vehicles) for the driver score, and with respect to ego velocity and side-to-side lateral distance for the cyclist score.

4.2.3.4 MDR

Among the overtaking scenarios, the Euclidean distance with the cyclist has an average of 2.64 meters with a standard deviation of 0.89.

Fig 4.8 shows the distribution of the MDR score.

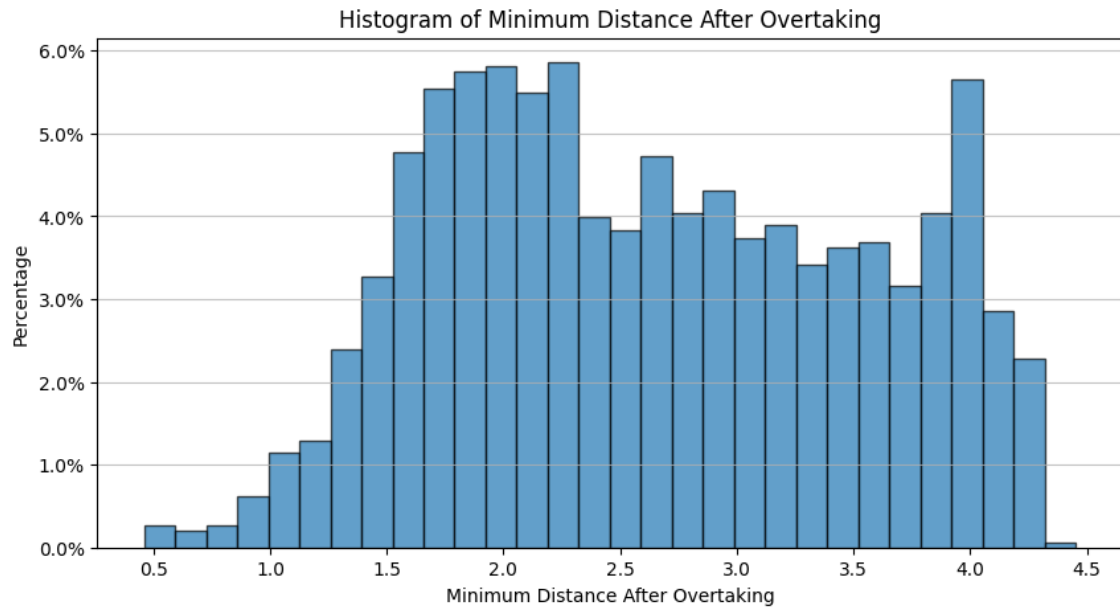


Figure 4.8: Distribution of Minimum Distance Returning values calculated across 1928 overtaking maneuvers.

Using 1 meter as the threshold, 1.3% of the cases are considered unsafe.

5

Discussion

5.1 Results from this thesis

In this section we discuss about the results from the last chapter.

5.1.1 Occurrence of Passing Types in Real World

From the large-scale dataset of over thousands of hours of naturalistic driving data that we have, meeting and drive-by cases are most common when the ego vehicle passes cyclists. Only around 4.5% of the passing events are identified as overtaking scenarios. About half (51.9%) of the cases are meeting cases, which is reasonable as the distinction between meeting and passing or overtaking is the direction (same or opposite) of the cyclist.

5.1.2 Validation from Manual Inspection

We performed manual inspection on all the 44 overtaking scenarios in the small batch dataset to validate the correctness of our implementation of scenario extractor and the four metrics.

Generally speaking, most of the scenarios contained a period when ego vehicle passed a cyclist, although some of them are more likely to be categorized into the drive-by scenario by human. Which means that the criteria for separating drive-by and overtaking cases can be further improved.

For the simple rule-based metrics LTRI, MDR and PRS, their results align well with the results from human annotators, as distance is a key input in these metrics and humans can also easily perceive risk levels from the distance information.

However for the perceived risk model, while the model and human agree well on the cases that are not risky, they may disagree on other risky cases. On the one hand, in these cases the model may even provide more accurate risk scores as human annotators are difficult to identify the exact speed and distance from video footage. On the other hand, human annotation incorporates all the available information from the scene, as real-world traffic is much more complicated than the controlled experiments, there are times that human annotation can provide more accurate risk scores.

Overall, the safety metrics demonstrate strong alignment with manual inspection, confirming their correctness.

5.1.3 Scoring Methods in Perceived Risk Score Model

We used the model weights from previous work by Rasch et al.[41] that used data containing perceived risk scores from field-track experiment and test-track experiment. During our research on this model, we identified abrupt score changes (for example sudden change from score 2 to 5 with a slight change in velocity) in the original scoring system that uses the score with maximum probability as the result. We altered the scoring system and resulted in the weighted scoring method.

From the comparison on the score difference (Fig 4.3) in the results chapter, both scoring methods give similar low scores on the safe cases when lateral distance is larger than 2 meters. But for other riskier cases the weighted scoring method usually return higher risk scores.

The weighted scoring method on the whole demonstrated improved consistency and smoother transitions between neighboring data points compared to the original method. For example, cases with similar driving conditions showed abrupt score jumps in the original method, while the weighted method produced more gradual and interpretable score changes. One potential drawback for the weighted method is that the average risk score is increased and there are more medium scores in the results, meaning that the modified method tends to return higher risk scores.

5.1.4 Results on Safety Metrics

LTRI was applied to both drive-by and overtaking scenarios in the phase/phases before passing phase. The results show that 10.7% of the drive-by cases were considered avoidable accident and 17.8% of the overtaking cases were considered avoidable accident. It suggests that overtaking maneuver is riskier than drive-by maneuver in terms of this metric.

VAMPD was also applied to both scenarios in the passing phase. The results showed that minimum lateral distance violations occurred in both drive-by and overtaking scenarios. In regions with MPD regulations, about one-third of the events failed to meet the required distance. In unregulated regions using the 1-meter assumption, around one-ninth of the events were considered unsafe.

PRS was applied in the passing phase in overtaking scenarios. We can confirm the clear trend from Fig 4.7 that the driver score is highly relevant with TTC with oncoming vehicle and ego velocity, the lower TTC and lower velocity, the higher driver's perceived risk score. We can also confirm that the cyclist score is highly relevant with lateral distance and ego velocity, the lower distance and higher velocity, the higher cyclist's perceived risk score.

MDR was applied in the returning phase in overtaking scenarios. It showed that most returning maneuvers maintained safe distances, though 1–2% fell below the deliberate 1-meter threshold. But we can also choose higher thresholds (for example 1.5 meters, which consequently identified 41 unsafe cases) to promote higher level of safety.

We can use these result to provide post-event feedback to the driver about each phase in the drive-by or overtaking. As well as to provide insights to user-based insurance if the results shows a trend, for example, if a driver tends to make sharp returns after overtaking cyclists, we can increase the insurance premium as the likelihood

for this driver having a rear-end crash with a cyclist is higher.

5.2 Special Cases

In this section we discuss some special cases were found that did not follow the usual patterns or caused by unexpected errors.

5.2.1 Overtaking a Group of Cyclists

According to Lopez et al.[28], overtaking a group of cyclists is a more complex and challenging scenario compared to overtaking a single cyclist. These situations often occur on rural or secondary roads, where road width may be limited and designated cycling infrastructure is typically absent. This situation requires experience and driving skills for the human drivers to perform a safe overtaking, requiring careful assessment of available space, timing, and the presence of oncoming traffic.

During visual inspection, we identified one such case that happened on a 2.5-meter wide two-lane road in the rural area of France, where the ego vehicle driver have been following a linearly aligned cycling group of 8 cyclists for over twenty seconds, allowing a few oncoming vehicles to pass, before initiating the overtaking. The driver kept at a speed of around 30 km/h when tailing the cycling group, and accelerated to up to 80 km/h during the overtake, showing a typical accelerative maneuver. The passing phase of overtaking the cyclist group lasted for 3.5 seconds, ego vehicle kept a lateral distance at around 1.8 to 1.9 meters, with velocity increasing from 55 km/h to 80 km/h, and has a time-to-collision with the oncoming vehicle of 5 seconds at the end of the passing phase.

The hard case here is that the sensors (and likely the human driver too) could not confidently identify the number and the length of the cyclist group. Since the cycling group rode in a straight line, one behind another, the sensors can only well identify the type and the position of the last cyclist, but not the other cyclists in front, especially for some sub-groups riding closely. The sensors can not give a confident result unless the ego vehicle is close enough to the specific cyclists.

In this case in France, the scenario extraction algorithm which we wrote only outputs the results regarding the last cyclist.. The remaining cyclists in the group were completely missed by the finder, likely due to low detection or classification confidence.

5.2.2 Misclassification of Objects

During our manual video review, we noticed two individual cases of misclassification in the sensor fusion, both happen at night.

One case takes place on nighttime urban roads, where parked cars on the opposite lane were identified as vehicles, and thus interpreted by our scenario extractor as oncoming vehicles. However, human driver should be able to understand that these were just parked cars and still considered it safe to overtake.

Another case also takes place at night, where a green mini-bus with large open windows was identified as the cyclist, causing it to wrongly identified by the scenario

extractor, and wrong dimensional calculation in lateral distance as we consider all cyclists has a fixed width.

This caused a mismatch between how humans and the model judged overtaking safety. As a result, some overtakes that were actually safe were labeled as risky by the model.

5.2.3 Overtaking on Curved Road

Some overtaking maneuvers happen on curved roads, both in the urban and the rural areas. For example, we have manually identified an overtaking scenario that happened on a curved mountain road with clear steering maneuver aiming to perform the overtake. However, in our current model, the analysis on overtaking scenarios is currently limited to scenarios that take place on straight roads. This is because the current detection of steering maneuver relies on signals like lateral distance change, lateral acceleration and steering wheel angle, which further tells apart passing and overtaking. But this could lead to wrongly-identified steering maneuver for data collected on curved roads, as there are no easy ways to identify the intention for the steering maneuver under different circumstances. This means that there can be some overtaking scenarios happen on curved road wrongly categorized into the drive-by scenarios. Furthermore, this could also lead to errors during phase separation.

To address this, future work could explore more robust methods for incorporating road curvature into scenario classification. For example, adding another sub-scenario in the scenario extractor, where the overtaking happens on the curved road, and adjust the metrics for detecting steering maneuver correspondingly. To do this, future work needs to be done to build more sophisticated models on identifying the driver's intention when steering (i.e. the driver steered in order to make a turn, to adjust the lane, to avoid obstacles, to overtake a leading object, etc). For such model, using machine learning based models may be a better solution due to the increased number of relevant variables like speed, road condition, road curvature, status of other road users, etc. And it may require intensive manual data annotation to build such classification model.

5.3 Limitations

In this section we discuss the limitations in our model, data source, sensor fusion process and other factors.

5.3.1 Rule-Based Safety Model

The rule-based safety models used in this study, such as LTRI, VAMPD, and MDR are easy to understand and explain. However, they also have some important limitations.

First, these models depend on pre-defined thresholds and rules. For example, in the VAMPD model, we use a default minimum lateral distance of 1 meter in areas without official MPD laws. While this makes the model simple to apply, it may not

match real-world driving habits or road conditions in different countries. This can lead to incorrect judgments in some cases.

Second, in some situations, such as noisy sensor data or unusual driving actions—the rules may not work well and could miss risky events or give wrong results.

Third, rule-based models are usually not flexible. They only work within the conditions we define. If a new or rare situation happens that doesn't fit the rules, the model may not be able to handle it correctly.

In short, rule-based models are simple and clear, but they may not perform well in complex or unexpected situations. In the future, combining rule-based and data-driven models might help balance clarity and flexibility.

5.3.2 Data Source and Reproducibility

One of the major limitations of this study lies in its reliance on proprietary data provided by Volvo Cars Corporation.

Due to the ethical and safety concerns associated with real-world overtaking scenarios involving vulnerable road users, such as cyclists, there are currently no publicly available datasets capturing these events in sufficient detail.[15] Conducting controlled experiments that involve vehicles overtaking cyclists at real driving speeds would pose significant risks and is generally considered ethically unacceptable. As a result, the collection of high-fidelity data in such scenarios is largely limited to large automotive manufacturers like VCC who have access to extensive test vehicles, internal system support, and quantitative naturalistic driving data.

In this study, all data used for data analysis and model evaluation is exclusively sourced from VCC's internal field datasets. These time-series datasets include multi-modal sensor information are collected from test vehicles in naturalistic driving environments. However, due to proprietary restrictions, data privacy regulations, and concerns over competitive advantage, these internal datasets are typically not shared publicly.

This poses a limitation in terms of reproducibility and external validation. Researchers outside of such industrial contexts may find it challenging to exactly reproduce or benchmark against our work, as they lack access to comparable datasets. Furthermore, even within industry, the diversity in sensor configurations and data logging practices across companies makes cross-validation difficult. While we strive to provide thorough documentation of our methods, the absence of a publicly accessible dataset inherently limits transparency and generalization in broader research contexts.

Some potential ways to address this issue are to either contribute anonymized, de-identified segments of the dataset to a shared research repository, or to develop simulation environments that can accurately replicate real-world overtaking events in a safe and reproducible manner. Such advancements would support broader academic collaboration and accelerate progress in this research area.

5.3.3 Sensor Fusion

Sensor fusion (SF) is the process of combining data from multiple sensors to produce an accurate and comprehensive understanding of the surrounding environment of the vehicle. This sensor-fused data is essential for various perception tasks, including object detection, tracking, and classification.

Logged data can be re-simulated using offline sensor fusion algorithms. This provides a visual platform the capability to represent the map location, ego object, and the traffic objects in certain scenarios allowing us to manually inspect the tracked surrounding traffic objects.

In our thesis work, we rely on the output from a sensor fusion algorithm which is available at VCC. We implemented the scenario extractor using the sensor-fused log data provided by VCC, and manually inspect the scenario in the visualization platform to manually analyze the scenario, identify hard cases and validate the modeling results.

While sensor fusion is a critical component of the perception system, it operates at a foundational level of the software stack and requires deep integration with vehicle hardware, calibration settings, and low-level algorithm design. As master thesis workers, we do not have access to modify or retrain the sensor fusion module itself, as it is maintained by specialized teams within the company and subject to strict quality assurance and validation procedures. Moreover, even if we have the access, developing or altering the sensor fusion pipeline would require extensive domain knowledge in areas such as real-time systems, sensor calibration, signal processing, and multi-sensor data fusion algorithms. These are the skills and expertise that go far beyond the scope and time constraints of a master's thesis. Thus, our focus in this thesis work, is rather to work on the outputs of the sensor fusion system and developing higher-level tools or models based on the available sensor-fused data.

5.3.3.1 Cyclist Dimension

There are inherent sensory limitations when it comes to accurately identifying the dimensions of a cyclist, primarily because cyclists do not have well-defined or rigid boundaries like motor vehicles. Unlike cars, which have standardized shapes and clear physical edges, cyclists vary in posture, clothing, and movement, and their outlines can be irregular or partially obscured, especially from the perspective of the cameras or the LiDARs. These factors make it challenging for sensors and perception algorithms to consistently detect the exact width and position of the cyclist, which can impact the accuracy of distance estimation and risk assessment during overtaking maneuvers. Addressing these limitations will require improved sensing techniques or advanced modeling approaches that can better account for the variability and ambiguity in cyclist profiles.

In this thesis work, we assume all cyclists are rectangular objects with a fixed width of 0.75 meters and uses the length information from the sensors.

5.3.3.2 Uncommon Type of Road User

In the sensor fusion results that we had access to, the classification of objects was limited to a finite set. Some types of objects such as e-scooters and snow cleaners do not have their own class.

During visual inspection on the camera footage and the visualized data, we have observed instances where a e-scooter was wrongly identified as pedestrian; a bicycle with a storage cart was misclassified as both a bicycle and a car; a tram was identified as a subsequent series of trucks. These errors in object type and classification lead to further issues, such as incorrect object width and length estimation, which further affects the accuracy for our model.

5.3.4 Other Factors

Some other factors also contribute to the risk severity of the overtaking event, however, they are not included in the sensor data, not processed, or cannot be quantified.

5.3.4.1 Lack on Environmental Information

During case studies about some ego-cyclist crash during overtaking maneuvers, sudden exposure to bright sunlight appears to have temporarily blinded the driver, contributing to the accident. This indicates that environmental factors can significantly influence driver behavior and pose threats to driver awareness and vehicle control, especially in situations where ego vehicle is intending to perform a cyclist overtake. However, such environmental information is not included or processed in the available data used for analysis.

The sensor fusion system focuses on detecting and tracking physical objects based on data from various sensors. It does not account for dynamic environmental conditions such as specific lighting condition and weather condition. As a result, factors like strong or sudden sunlight glare, which can impair both human drivers and camera-based perception systems, are not reflected in the sensor-fused data and are difficult to identify or quantify reliably from the available sensor data. Furthermore, inferring the presence and impact of strong sunlight from camera footage alone may be technically complex. The visibility of glare depends on many variables, including sun angle, time of day, road direction, and sensor characteristics. These variables are not explicitly logged or modeled in the system we have.

Because of these limitations, critical environmental influences like sunlight glare or slippery icy road currently could only be identified through manual annotation, which restricts the scalability and objectivity of the results. Addressing this problem in the future would require enhanced sensor capabilities and better detection algorithms on such scenarios.

5.3.4.2 Cyclist Unsafe Behavior

In one particularly notable case, we observed that the cyclist was riding with his both hands off the handlebar while the ego vehicle was approaching at a close lateral distance on a relatively narrow rural road. This situation was highly critical and

dangerous for the cyclist as riding without hands on the handlebar significantly reduces the cyclist’s ability to maintain balance and the ability to react quickly to sudden changes in the environment, making them more vulnerable to instability or swerving, especially when a vehicle is trying to perform overtaking. This scenario also raises the driver’s perceived danger in real life, as the driver needs to stay vigilant in case of any unexpected cyclist movement.

Ideally, in such real-life scenarios, the driver of the ego vehicle should maintain a significantly larger lateral distance from the cyclist to account for the increased unpredictability and risk. However, the current sensor fusion system and visualization system do not process or interpret the cyclist’s specific posture, such as whether their hands are on the handlebars. As a result, this important behavioral cue is not represented in the sensor data, as the steering status of the cyclist is not processed, which makes this case a normal overtaking. This case would be categorized as safe in our model, despite it is critically unsafe in real life for human observers.

5.4 Future Work

Based on our thesis work, there are several directions that could be explored in the future to improve and extend this study.

5.4.1 Scenario Extractor

5.4.1.1 Steering Maneuver

Detecting steering maneuver and its exact start time was a major difficulty during our thesis progress and it remains to be further investigated. We have several metrics for this detection: lateral acceleration, yaw rate, steering wheel angle and lateral distance change. The first three metrics are highly correlated, and all of these metrics are sensitive to road curvature. Currently we are using maximum steering wheel angle to indicate whether and approximately when the steering maneuver happens. We chose this value as this is a direct driver input that best indicates the driver’s intention. However, as this paper suggests[39], in the future these values can be further investigated to more accurately detect the presence and the starting time of the steering maneuver.

5.4.1.2 Overtaking Other Road Users

Although our model was designed specifically for overtaking cyclists, the core idea can be extended to other types of road users. In real traffic, drivers often overtake not only cyclists but also pedestrians, e-scooter users, wheelchairs, agricultural vehicles, or even delivery robots on the road edge.

The method of detecting risky overtaking behavior — by analyzing distance, speed, and relative movement — can be adapted for these other scenarios. Each type of road user may require different threshold values and safety indicators, but the overall framework remains applicable.

This opens up possibilities for future research to build a more generalized overtaking risk evaluation system that supports a wider range of vulnerable or slow-moving road users. Expanding the dataset and fine-tuning the model for new categories would be the next logical step.

5.4.2 Models

5.4.2.1 Squeezing Strategy

According to the study by Rasch et al.[40], and our own manual inspection of the data, apart from the flying and accelerative overtaking strategies, we identified a third overtaking strategy, namely 'squeezing'. This strategy can be identified by the driver's maneuver to pass between the cyclist and the oncoming vehicle, which means there can be longitudinal overlap for the ego vehicle, the cyclist and the oncoming vehicle in the passing phase. This maneuver presents unique challenges and safety concerns, as it requires a high level of precision and judgment from the driver.

In addition to the factors already considered in overtaking scenarios, such as ego velocity and lateral distance to the cyclist, the 'squeeze through' strategy requires a more comprehensive evaluation of factors like the lateral distance between the ego vehicle and the oncoming vehicle, as well as the relative speed between the ego vehicle and the oncoming vehicle. These factors are critical in determining whether this overtaking strategy can be executed safely without risking a collision.

To better understand and mitigate the risks associated with this type of overtaking strategy, it will be important to develop another safety model during the passing phase that incorporates these additional variables.

As a future work, such model could provide valuable insights into the limits of safe maneuvering in these complex situations and guide future vehicle automation systems in making safer overtaking decisions.

5.4.2.2 Cyclist Velocity

Cyclist velocity may also play a significant role in determining the safety of overtaking maneuvers. When a cyclist is traveling at a lower speed, their bicycle may be less stable according to this stability analysis[45], as lower speeds can reduce the rider's ability to react to sudden changes in the environment, particularly in the presence of overtaking vehicles. On the other hand, higher cyclist speed presents an increased danger, as higher cyclist speed reduce the available reaction time for the driver of the ego vehicle to access and execute a safe overtaking maneuver, especially when the cyclist were forced to make a sharp turn. Thus, the cyclist's velocity acts as a dual factor: lower speeds compromise stability and control, while higher speeds pose extra threat for safe overtaking, both contributing to the overall risk of accidents in these scenarios.

For future work, cyclist velocity could be used to train perception models, improving the system's ability to predict cyclist behavior during overtaking maneuvers. By incorporating this factor, the model can better assess both driver's and cyclist's perceived risks, and enhance decision-making for safer interactions with cyclists.

5.4.2.3 Type of Oncoming Vehicle

Previous review[43] summarizes that buses keep the minimum passing lateral distance when overtaking cyclists, while trucks keep the largest lateral distance comparing to all types of common motor vehicles. This indicates that the ego vehicle driver of different types of vehicles are willing to take different levels of risks when overtaking cyclist. From this and together with our own experience, the type of the oncoming vehicles may have an influence on driver's decision-making and perceived level of safety during cyclist overtaking maneuvers. For example, the ego vehicle driver may perceive a higher risk in the presence of an oncoming truck and adjust their overtaking strategy accordingly during the approaching phase, or the driver may choose to finish the overtaking more quickly with an increased speed, a quickened returning maneuver, and a reduced Euclidean distance with the cyclist during returning phase in order to avoid head-on crash with the oncoming truck, which on the other hand, resulting in a sharp drop in perceived safety for the cyclist.

To further investigate this potentially decision-influencing factor, we should extract the type of the oncoming vehicle, and consider it in the modeling process.

6

Conclusion

This thesis explored how to identify and analyze the way vehicles overtake cyclists using real-world driving data collected from Volvo test vehicles. By developing a scenario extraction tool and applying it to over thousands of naturalistic driving streams, we were able to detect sufficient overtaking scenarios.

In this thesis work, we first conducted data analysis on the time-series log data, followed by the implementation of a new scenario finder within VCC’s internal scenario extractor tool to find new scenarios. This new scenario finder allows us to identify scenarios in which the ego vehicle drives by or overtakes a cyclist, with the potential presence of multiple oncoming vehicles. Furthermore, we segmented each scenario into distinct phases: four for overtaking scenario (**approaching, steering away, passing, and steering back**), and three for drive-by scenario (**approaching, passing, and departing**). At each phase, we derived relevant data from signals, such as time to danger, lateral distance, vehicle velocity, time-to-collision with oncoming vehicles, minimum returning distance. These ways to segment scenarios helped us understand not just when overtaking happens, but also how safety-critical situations occur.

To address the complexity of overtaking scenario, we defined four safety metrics (VAMPD, PRS, LTRI, MDR) to evaluate the risk. They are applied across different phases within two categories (Drive-by and Overtaking), depending on their relevance. Variable-Adjusted Minimum Passing Distance (VAMPD) is based on specific traffic rules and we made a point that the minimum passing distance is not only adjusted by ego speed but also the road type (urban or rural). This safety metric was applied in the passing phases both drive-by and overtaking scenario. Perceived Risk Score (PRS) is a safety metric with a continuous risk score for both the driver and the cyclist perspective. This safety metric was applied in the passing phase of overtaking scenario. LD-TTD Risk Index (LTRI) is a safety metric that combines time to danger with lateral distance. This safety metric was applied in the approaching phase of the drive-by scenario and the steering away phase of the overtaking scenario. Minimum Distance Returning (MDR) is a safety metric specific for the steering back phase of overtaking scenario by analyzing Euclidean distance between ego vehicle and cyclist.

In addition to implementing and evaluating safety metrics across different overtaking phases and scenarios, we also identified a third overtaking strategy — *squeezing*. This maneuver, where the driver passes between a cyclist and an oncoming vehicle, raises specific safety concerns due to the close proximity of all parties involved. Its occurrence highlights the need for further research on high-risk overtaking behaviors in real-world driving.

During the project, we also faced several challenges, especially when working with real-world driving data. Unlike experimental data, naturalistic data is more complex: roads can be curved, sensors may not always be accurate, and cyclists behavior can vary. It was particularly difficult to break the overtaking events into clear phases when the signals were noisy or missing. In addition, our current analysis does not take into account the type or speed of the cyclist, or the type of oncoming vehicles. These factors may also affect overtaking safety and should be explored in future work.

While our safety model works well and the evaluation results of the model show high consistency with manual inspection, validating its usability and accuracy, it still struggles to handle the unpredictability of the real world. We also looked at some of the more complex cases like overtaking a group of cyclists, sensor unreliability at night, overtaking maneuvers happening on curved roads and discussed how current sensors and models struggle to handle these less straightforward situations.

To summarize, the work presented in this thesis helps improve the understanding of how drivers overtake cyclists and may be valuable for providing post-event feedback, supporting user-based insurance systems, and ultimately promoting road safety.

What our thesis work contributes

- A practical tool for extracting cyclist-related passing scenarios from massive naturalistic driving data.
- A detailed way to break down overtaking into phases and study driver behavior at each phase.
- A first step toward classifying overtaking strategies in natural driving conditions.
- Improvement and comparison on an exiting perceived risk model.
- A data analysis on naturalistic driving data.
- A safety model consist of four metrics focusing on different phases.
- Identifying edge cases in real-world cyclist overtaking scenarios.

Bibliography

- [1] Mina Alibeigi, William Ljungbergh, Adam Tonderski, Georg Hess, Adam Lilja, Carl Lindstrom, Daria Motorniuk, Junsheng Fu, Jenny Widahl, and Christoffer Petersson. Zenseact open dataset: A large-scale and diverse multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.
- [2] Subramanian Arumugam and R Bhargavi. A survey on driving behavior analysis in usage based insurance using big data. *Journal of Big Data*, 6(1):1–21, 2019.
- [3] Jackie Ayoub, Zifei Wang, Meitang Li, Huizhong Guo, Rini Sherony, Shan Bao, and Feng Zhou. Cause-and-effect analysis of adas: A comparison study between literature review and complaint data. In *Proceedings of the 14th international conference on automotive user interfaces and interactive vehicular applications*, pages 139–149, 2022.
- [4] Ben Beck, Derek Chong, Jake Olivier, Monica Perkins, Anthony Tsay, Adam Rushford, Lingxiao Li, Peter Cameron, Richard Fry, and Marilyn Johnson. How much space do drivers provide when passing cyclists? understanding the impact of motor vehicle and infrastructure characteristics on passing distance. *Accident Analysis & Prevention*, 128:253–260, 2019.
- [5] Siri Hegna Berge, Joost de Winter, Diane Cleij, and Marjan P. Hagenzieker. Triangulating the future: Developing scenarios of cyclist-automated vehicle interactions from literature, expert perspectives, and survey data. *Transportation Research Interdisciplinary Perspectives*, 21:100986, 2024.
- [6] Christian-Nils Boda, Marco Dozza, Pablo Puente Guillen, Prateek Thalya, Leila Jaber, and Nils Lubbe. Modelling discomfort: How do drivers feel when cyclists cross their path? *Accident Analysis & Prevention*, 146:105550, 2020.
- [7] Tom Brijs, Filomena Mauriello, Alfonso Montella, Francesco Galante, Kris Brijs, and Veerle Ross. Studying the effects of an advanced driver-assistance system to improve safety of cyclists overtaking. *Accident Analysis & Prevention*, 174:106763, 2022.
- [8] Ralph Buehler and John Pucher. *Cycling for Sustainable Cities*. MIT Press, Cambridge, MA, USA, 2021.
- [9] Bundesministerium für Digitales und Verkehr. Straßenverkehrs-Ordnung (StVO). https://www.gesetze-im-internet.de/stvo_2013/, 2023. Accessed: 2025-05-12.
- [10] Shengdi Chen, Qingwen Xue, Xiaochen Zhao, Yingying Xing, and Jian John Lu. Risky driving behavior recognition based on vehicle trajectory. *International Journal of Environmental Research and Public Health*, 18(23):12373, 2021.

- [11] Qiang Cheng, Hao Li, Yifan Yang, Jun Ling, and Xiaotong Huang. Towards efficient risky driving detection: A benchmark and a semi-supervised model. *Sensors*, 24(5):1386, 2024.
- [12] William Cubbin, Kjell van Paridon, Helen Keyes, and Matthew Timmis. Close passes caught on camera—how knowledge and behavioural norms relate to perceptions of liability when cars overtake cyclists. *Transportation research part F: traffic psychology and behaviour*, 100:308–322, 2024.
- [13] Dirección General de Tráfico. Spain Introduces New 2025 Cycling Regulations to Protect Vulnerable Road Users, 2024. Accessed: 2025-05-12.
- [14] Marco Dozza, Ron Schindler, Giulio Bianchi-Piccinini, and Johan Karlsson. How do drivers overtake cyclists? *Accident Analysis & Prevention*, 88:29–36, 2016.
- [15] Inês Dutra and Joaquim Fonseca. Data and knowledge for overtaking scenarios in autonomous driving. *Journal of Autonomous Vehicles and Systems OCTOBER*, 2:041006–1, 2022.
- [16] Thodoris Garefalakis, Eva Michelaraki, Stella Roussou, Christos Katrakazas, Tom Brijs, and George Yannis. Predicting risky driving behavior with classification algorithms: results from a large-scale field-trial and simulator experiment. *European Transport Research Review*, 16(65), 2024.
- [17] GOV.UK. The Highway Code - Using the Road (159 to 203), 2025. Accessed: 2025-05-12.
- [18] Kristiann C Heesch, Amy Schramm, Ashim Kumar Debnath, and Narelle Harworth. Cyclists’ perceptions of motorist harassment pre-to post-trial of the minimum passing distance road rule amendment in queensland, australia. *Health promotion journal of Australia*, 28(3):247–250, 2017.
- [19] Geertje Hegeman, Andreas Tapani, and Serge Hoogendoorn. Overtaking assistant assessment using traffic simulation. *Transportation Research Part C: Emerging Technologies*, 17(6):617–630, 2009.
- [20] Andres Henao and Philippe Apparicio. Dangerous overtaking of cyclists in montréal. *Safety*, 8(1):16, 2022.
- [21] Xianbiao Hu, Xiaoyu Zhu, Yu-Luen Ma, Yi-Chang Chiu, and Qing Tang. Advancing usage-based insurance—a contextual driving risk modelling and analysis approach. *IET Intelligent Transport Systems*, 13(3):453–460, 2019.
- [22] IEEE P2020 Working Group. Ieee p2020 automotive imaging white paper, 2018.
- [23] Citizens Information. Laws on cycling in ireland, 2024. Accessed: 2025-05-12.
- [24] I Isaksson-Hellman and J Werneke. Detailed description of bicycle and passenger car collisions based on insurance claims. *Safety science*, 92:330–337, 2017.
- [25] Joon-Ki Kim, Sungyop Kim, Gudmundur F Ulfarsson, and Luis A Porrello. Bicyclist injury severities in bicycle–motor vehicle accidents. *Accident Analysis & Prevention*, 39(2):238–251, 2007.
- [26] Long Liu, Zhelong Wang, and Sen Qiu. Driving behavior tracking and recognition based on multisensors data fusion. *IEEE Access*, 8:194671–194681, 2020.

-
- [27] Carlos Llorca, Antonio Angel-Domenech, Fernando Agustin-Gomez, and Alfredo Garcia. Motor vehicles overtaking cyclists on two-lane rural roads: Analysis on speed and lateral clearance. *Safety science*, 92:302–310, 2017.
- [28] Griselda Lopez, Ana María Pérez-Zuriaga, Sara Moll, and Alfredo García. Analysis of overtaking maneuvers to cycling groups on two-lane rural roads using objective and subjective risk. *Transportation research record*, 2674(7):148–160, 2020.
- [29] Shane B. McLaughlin, Jonathan M. Hankey, Thomas A. Dingus, and Sheila G. Klauer. Development of an few algorithm evaluation methodology with evaluation of three alert algorithms. Final Report DOT HS 811 145, National Highway Traffic Safety Administration, 2009.
- [30] Sara Moll, Griselda López, Alexander Rasch, Marco Dozza, and Alfredo Garcia. Modelling duration of car-bicycles overtaking manoeuvres on two-lane rural roads using naturalistic data. *Accident Analysis & Prevention*, 160:106317, 2021.
- [31] National Highway Traffic Safety Administration. Motorist Passing Bicyclist Laws, 2025. Accessed: 2025-05-12.
- [32] New South Wales Government. Road Rules 2014 - Regulation 144-1: Keeping a Safe Distance When Passing Bicycle Rider. https://www5.austlii.edu.au/au/legis/nsw/consol_reg/rr2014104/s144.1.html, 2014. Accessed: 2025-05-12.
- [33] NHTSA. Vehicle owner’s complaint database, 2011. Accessed: 2025-05-21.
- [34] World Health Organization. *Global Status Report on Road Safety*. World Health Organization, Geneva, Switzerland, 2018.
- [35] World Health Organization. *Global status report on road safety 2018*. World Health Organization, 2019.
- [36] World Health Organization. Road traffic injuries, 2023. Accessed: 2024-11-28.
- [37] Tatsuaki Osafune, Toshimitsu Takahashi, Noboru Kiyama, Tsuneo Sobue, Hirozumi Yamaguchi, and Teruo Higashino. Analysis of accident risks from driving behaviors. *Journal of Advanced Transportation*, 15:192–202, 2016.
- [38] Alexander Rasch, Christian-Nils Boda, Prateek Thalya, Tobias Aderum, Alessia Knauss, and Marco Dozza. How do oncoming traffic and cyclist lane position influence cyclist overtaking by drivers? *Accident Analysis & Prevention*, 142:105569, 2020.
- [39] Alexander Rasch and Marco Dozza. Modeling drivers’ strategy when overtaking cyclists in the presence of oncoming traffic. *IEEE Transactions on Intelligent Transportation Systems*, 23(3):2180–2189, 2022.
- [40] Alexander Rasch, Carol Flannagan, and Marco Dozza. When is it safe to complete an overtaking maneuver? modeling drivers’ decision to return after passing a cyclist. *IEEE Transactions on Intelligent Transportation Systems*, 2024.
- [41] Alexander Rasch, Sara Moll, Griselda López, Alfredo García, and Marco Dozza. Drivers’ and cyclists’ safety perceptions in overtaking maneuvers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 84:165–176, 2022.

- [42] Alexander Rasch, Yury Tarakanov, Gustav Tellwe, and Marco Dozza. Drivers passing cyclists: How does sight distance affect safety? results from a naturalistic study. *Journal of safety research*, 87:76–85, 2023.
- [43] Elisabeth Rubie, Narelle Haworth, Divera Twisk, and Naohide Yamamoto. Influences on lateral passing distance when motor vehicles overtake bicycles: A systematic literature review. *Transport reviews*, 40(6):754–773, 2020.
- [44] Elisabeth Rubie, Narelle Haworth, and Naohide Yamamoto. Passing distance, speed and perceived risks to the cyclist and driver in passing events. *Journal of safety research*, 87:86–95, 2023.
- [45] Arend L Schwab and Jaap P Meijaard. A review on bicycle dynamics and rider control. *Vehicle system dynamics*, 51(7):1059–1090, 2013.
- [46] Ke Wang, Qingwen Xue, and Jian John Lu. Risky driver recognition with class imbalance data and automated machine learning framework. *International Journal of Environmental Research and Public Health*, 18(14):7534, 2021.
- [47] De Jong Yeong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, 21(6):2140, 2021.

A

Appendix 1

Table A.1: Key variables and their definitions used in the study

Variable	Unit	Description
LD	m	Lateral Distance between the ego vehicle and the cyclist
ED	m	Euclidean Distance
V_{ego}	km/h	Speed of the ego vehicle during overtaking
V_{on}	km/h	Speed of the oncoming vehicle during overtaking
TTC	s	Time To Collision between two vehicles
TTC_{on}	s	Time To Collision with oncoming vehicle at passing moment
TTD	s	Time To Danger with cyclist
a_{ego}	m/s^2	Longitudinal acceleration of the ego vehicle
d_{cyclist}	m	Distance from ego vehicle to the cyclist
θ	degrees	Steering angle of the ego vehicle
Δt	s	Time difference between overtaking phases
s	m	Longitudinal displacement of the vehicle
w_{veh}	m	Width of the ego vehicle
w_{cyc}	m	Width of the cyclist
w_{lane}	m	Width of the lane
N_{events}	count	Number of overtaking events recorded
f_{FPS}	Hz	Frame rate of the video data

B

Appendix 2

Table B.1: Minimum Passing Distance (MPD) laws for overtaking cyclists in different countries and regions

Country / Region	MPD (Minimum Passing Distance)	Is MPD Affected by Speed Limit?
Australia (NSW) [32]	1 m (≤ 60 km/h), 1.5 m (> 60 km/h)	Yes
Germany [9]	1.5 m (urban), 2 m (non-urban)	No
Spain [13]	1.5 m	Yes (Reduce speed by 20 km/h below limit)
Ireland [23]	1 m (≤ 50 km/h), 1.5 m (> 50 km/h)	Yes
USA (varies by state) [31]	Most states require ≤ 3 ft (≈ 0.91 m) ; PA = 4 ft, SD = 3–6 ft	Dependence (varies by state)
United Kingdom [17]	1.5 m (≤ 30 mph ≈ 48 km/h), More space required (> 30 mph)	Yes (more space at higher speeds)

Note on USA Law: As of 2021, the majority of U.S. states—35 states and the District of Columbia—had implemented laws requiring drivers to maintain a minimum lateral distance of at least 3 feet when overtaking cyclists [31]. Some states have stricter requirements: for example, Pennsylvania and New Jersey mandate a minimum of 4 feet, while South Dakota requires 3 feet on roads with speed limits of 35 mph or less, and 6 feet on faster roads. In North Carolina, a minimum of 2 feet is required, and drivers may pass cyclists in no-passing zones if they maintain at least a 4-foot buffer. Additionally, states such as Delaware, Kentucky, Nevada, Oklahoma,

and Washington mandate lane changes to pass cyclists on multi-lane roads. Eight other states follow more general standards, requiring a “safe distance,” though the term is not precisely defined in those jurisdictions. In any other countries without specific traffic rules for overtaking cyclists, we assume that it is risky once the lateral distance is less than 1 meter, as this does not provide adequate space for the safety of vulnerable road users.

C

Appendix 3

Table C.1: Output variables from *EgoPassingCyclistFinder* in Scenario Extractor – General Information

Output Name	Definition (Signals Used)	Usage / Note
object_class	From object signal [str / {BICYCLE, UNKNOWN}]	To filter cyclists
obj_width	From object signal [float / m]	Always 0.5 m, but should use 0.75 m in later processes
obj_length	From object signal transient length, take the median value of 2 seconds before passing [float / m]	Calculate object length in the period when sensors have better accuracy

Table C.3: Output Variables – Approaching & Steering Away Phase (Pre Passing Phase)

Output Name	Definition (Signals Used)	Usage / Note
max_road_curvature	From ego signal [float / rad]	Filter straight roads for steering detection
max_steering_angle _steeraway_abs	From ego signal [float / rad]	Detect steering maneuver
infer_steeraway_start	0.5 seconds before steering wheel angle reaches maximum in pre phase [float / seconds]	Detect if steering maneuver happened during pre phase
max_warning_score	Warning score in paper [int]	Indicator to avoid front-rear crash with the cyclist
max_lat_acc_pre_abs	From ego signal [float / m·s ⁻²]	Detect if steering maneuver happened during pre phase
min_long_acc_pre	From ego signal [float / m·s ⁻²]	Identify overtaking strategy (accelerative/flying)
max_long_acc_pre	From ego signal [float / m·s ⁻²]	Identify overtaking strategy (accelerative/flying)
lat_dist_delta _steeraway_abs	Max lateral distance change 2 seconds before max steering to passing start [float / m]	Detect if steering maneuver happened during pre phase
min_euclidean _distance_pre	Euclidean distance to the cyclist box [float / m]	Not used?
oncoming_present _steeraway	Whether oncoming object present [int / {0, 1}]	During steer away phase
min_ttc_steeraway	TTC with closest oncoming object during pre phase [float / s]	To-do: change to steer away phase?

DEPARTMENT OF MECHANICS AND MARITIME SCIENCES

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

www.chalmers.se



CHALMERS
UNIVERSITY OF TECHNOLOGY