Deep Learning for Abnormal Driving Behaviour Detection

T&P, CiTG
MSc Thesis (CIE5060-09)

to obtain the degree of
Master of Science Civil Engineering – Transport and Planning
at the Delft University of Technology
to be defended publicly on 30th June 2023.

Zhang Lanxin
5119863

Committee Members

Chair Supervisor	Prof.dr.ir. B. (Bart) van Arem
Daily Supervisor	Dr. ir. H. (Haneen) Farah
Daily Supervisor	Y. (Yongqi) Dong
External Supervisor	Dr. A. (Arkady) Zgonnikov

Preface

Road traffic safety is a pressing global concern, with millions of yearly fatalities and injuries. This study aims to address the detection of abnormal driving behaviour and proposes novel approaches utilizing machine learning techniques. The research was conducted in the context of the Civil Engineering Department TU Delft, under the supervision of Prof.dr.ir. B. (Bart) van Arem, the Chair Supervisor, and Dr. ir. H. (Haneen) Farah and Y. (Yongqi) Dong as the Daily Supervisors. Dr A. (Arkady) Zgonnikov contributed as the External Supervisor.

In traditional supervised approaches, limitations arise due to the labelled abnormal driving data requirement. This research explores and develops semi-supervised machine learning models to overcome this challenge. These models leverage the power of machine learning for abnormal driving behaviour detection, offering a data-driven approach that adapts to different scenarios and captures subtle patterns. Moreover, their scalability enables efficient analysis of large datasets, accurately identifying abnormal driving behaviour and providing valuable insights for enhancing road safety measures.

I would also like to extend a special thank you to my parents for their unwavering support and encouragement throughout my academic journey. Their love, sacrifices, and belief in my abilities have been instrumental in my success.

By the way, I have learned much during my time at TU Delft. It is a highly rigorous academic institution known for its commitment to excellence in research and education. Being at TU Delft has provided me with abundant academic resources and an environment conducive to learning. The emphasis on academic rigour and the integration of theory and practical applications has profoundly impacted my research work. I am grateful for the learning experience at TU Delft and will continue to apply the knowledge gained to my future research and practical endeavours.

Zhang lanxin
2023/6/25

Abstract

Road traffic safety is a pressing global concern, with millions of yearly fatalities and injuries. This study aims to address the detection of abnormal driving behaviour. Traditional supervised approaches face limitations due to the need for labelled abnormal driving data. To overcome this challenge, semi-supervised machine learning models are explored and developed in this research.

Machine learning is utilized for abnormal driving behaviour detection because it offers a data-driven approach that adapts to different scenarios and captures subtle patterns. Furthermore, its scalability allows for efficient analysis of large datasets, leading to accurate identification of abnormal driving behaviour and valuable insights for enhancing road safety measures. Most existing machine learning (ML) based abnormal driving detectors rely on (fully) supervised ML methods, which require substantial labelled data. However, in the real world, labels are only sometimes available, and labelling large amounts of data is tedious. Thus, there is a need to employ unsupervised or semi-supervised methods to make the detection process more feasible and efficient. Luckily, it is possible with the advent of deep neural networks, especially autoencoder-based ones. This thesis develops and compares three ML methods: supervised (e.g. XGBoost and Random Forest), unsupervised ML (e.g. Isolation Forest and Robust Covariance), and semi-supervised ML (Hierarchical Extreme Learning Machines). Comparison results show that the semi-supervised deep learning model outperforms unsupervised methods exhibiting higher prediction accuracy and delivering acceptable results compared to the fully supervised models.

Moreover, previous ML-based approaches predominantly utilize basic car motion features (such as velocity and acceleration) to label and predict abnormal driving behaviours. In contrast, this thesis introduces Surrogate Measures of Safety (SMOS) as features for ML models to identify abnormal driving behaviour.

The results indicate that the supervised model performs best under the same conditions. However, relying on a large amount of labelled data in supervised models can pose challenges in real-life scenarios or when dealing with massive datasets. The study highlights the significance of Surrogate Measures of Safety (SMOS) and demonstrates the potential of HELM in effectively identifying abnormal driving behaviour. The introduction of SMOS significantly improves the performance of both unsupervised and semi-supervised models. The unsupervised model shows the most substantial improvement, increasing accuracy from less than 50% to over 90%.

While the Isolation Forest and Robust Covariance models fail to detect abnormal driving behaviour without including SMOS, the semi-supervised HELM model exhibits

promising results even without SMOS. However, further research is necessary to address limitations and enhance the findings. While valuable, the current dataset used in this study may only encompass some types of abnormal driving behaviour. Future research should incorporate a more diverse dataset that covers a broader range of abnormal driving behaviours. The analysis should include multiple SMOS features, such as Post Encroachment Time (PET), to comprehensively understand abnormal driving behaviour and improve safety measures.

Keywords: road traffic safety, abnormal driving behaviour, machine learning, unsupervised learning, semi-supervised learning, surrogate measures of safety (SMOS), hierarchical extreme learning machines (HELM).

Table of content

Preface......1
Abstract......2
Table of content......4
Table List......6
Figure List......7
1 Introduction......8
1.1 Problem statement......8
1.2 Aim of research......9
1.3 Societal impact......10
1.4 Scope of research......11
1.5 Thesis outline......12
2 Literature review......13
2.1 Anomaly detection research overview......13
2.2 Abnormal driving detection research......14
2.2.1 Data collection methods......14
2.2.2 Abnormal driving behaviour......16
2.2.3 Traditional featuress used in ML-based methods......17
2.2.4 Surrogate measures of safety used in ML-based methods......18
2.2.5 Machine learning model......19
2.3 Conclusion......23
3 Research Gap and Research Questions......24
3.1 Research gaps......24
3.2 Research Questions......24
4 Data......26
4.1 Dataset requirements......26
4.2 Dataset description......26
4.2 Abnormal driving behaviour in the dataset......28
4.2.1 Signalized Intersection......29
4.2.2 Expressway......31
4.2.3 Non-signalized Intersection......36
4.3 Conclusion......37
5 Methodology......38
5.1 ML models......38
5.1.1 Supervised ML......38
5.1.2 Unsupervised ML......40
5.1.3 Semi-supervised......42
5.2 Ablation Study Regarding Features......44
5.2.1 Features Overview......44
5.2.2 Experiment Design......45

5.3 Model performance measures 47
5.4 Conclusion 48
6 Results, analysis and discussion 49
6.1 Comparison result for ML models 49
6.2 Result for Ablation Study Regarding Features 52
6.3 Conclusion 55
7 Conclusions and recommendations..58
7.1 Conclusions..58
7.2 Reflection..58
7.3 Recommendation 59
8 Reference 61

Table List

Table 1 Dataset overview........16
Table 2 Different definition of Abnormal Driving Behaviour by Chen et al., 2015; Kim et al., 2016........17
Table 3 Traditional features overview........18
Table 4 Overview of Machine Learning Applied for Abnormal Driving Detection........22
Table 5 Example of dataset........28
Table 6 Data collection information at Alafaya........29
Table 7 Data collection information at Expressway A........31
Table 8 Lane changing behaviour at Expressway A........32
Table 9 Lateral acceleration........33
Table 10 Data collection information at University McCulloch........37
Table 11 Post Encroachment Time........37
Table 12 Abnormal driving behaviour summary in this thesis........37
Table 13 Input features in different settings........46
Table 14 Confusion Matrix and the Corresponding Array Representation........47
Table 15 Comparison with different models........50
Table 16 Comparison results under different set........53
Table 17 Input features in different scenario........54
Table 18 Comparison with different models........56

Figure List

Figure 1 Anomaly detection....................14
Figure 2 Abnormal driving behaviour(Chen et al., 2015)....................17
Figure 3 Example of how supervised learning works (https://neurospace.io/blog/2020/08/what-is-supervised-learning) 20
Figure 4 Example of Unsupervised Learning 21
Figure 5 Example of Semi-supervised Learning 22
Figure 6 Vehicle bounding box feature description 27
Figure 7 Relationship between behaviour, indicators and location (drawn by *Zhang lanxin*) 29
Figure 8 Alafaya University Signalized Intersection....................29
Figure 9 Rapid acceleration and Emergency braking (Jia et al., 2020) 30
Figure 10 Extreme acceleration and deceleration points distribution at different speeds (red=normal/orange=abnormal) 31
Figure 11 Expressway A 31
Figure 12 Proportional distribution of lane changing times 32
Figure 13 Rapid lane changing (Jia et al., 2020) 33
Figure 14 Lateral Acceleration Distribution 34
Figure 15 Extreme lateral acceleration and deceleration points distribution at different speeds (red=normal/orange=abnormal) 34
Figure 16 Close Lane changing (Zheng et al., 2022)....................35
Figure 17 Distance distribution in Expressway A....................35
Figure 18 Extreme distance points distribution at different CarID (green<0.5m/0.5≤red≤1.0m/blue>1.0m)....................36
Figure 19 McCulloch Seminole Non-Signalized Intersection....................36
Figure 20 HELM Framework (Drawn by *Zhang Lanxin*) 44
Figure 21 2DTTC scenario (Xing et al., 2019)....................45
Figure 22 Supervised machine learning result....................49
Figure 23 Unsupervised machine learning result....................50
Figure 24 Semi-supervised machine learning result....................50
Figure 25 Robust Covariance performance under Set 1,2, and 3 53
Figure 26 Isolation Forest performance under Set 1,2, and 3....................53
Figure 27 HELM performance under Set 1,2, and 3....................53
Figure 28 Supervised machine learning result....................55
Figure 29 Unsupervised machine learning result....................55
Figure 30 Semi-supervised machine learning result....................56

1 Introduction

1.1 Problem statement

Road traffic safety has become a growing concern worldwide (Gerónimo et al., 2010). World Health Organization released data that approximately 1.35 million people died in car crashes in 2018 (World Health Organization, 2018). In addition, more than 30 million people suffered non-fatal injuries, many disabled. Traffic accidents also caused considerable economic loss to individuals, their families, and nations. In many countries, the financial loss is as much as 3% of their gross domestic product (World Health Organization, 2018). In 92.9% of accidents, human is a contributing factor (Saiprasert & Pattara-Atikom, 2013). Therefore, it is necessary to understand how driver behaviour contributes to unsafe situations and alleviate the accidents caused by abnormal human driving behaviours.

Driving behaviour is a broad concept that can function from many variables and factors, including driving performance, environmental awareness, willingness to take risks, and reasoning abilities(Mohammadnazar et al., 2021a). Driving style can substantially impact mobility, safety, energy consumption, and vehicle emissions (Mohammadnazar et al., 2021b). Abnormal driving behaviour is usually defined as actions that deviate from normal or safe driving. It involves engaging in behaviours that put oneself, passengers, and other road users at risk. Some examples of abnormal driving behaviour include(Academic et al.):

Excessive Speeding: Driving well above the posted speed limits or too fast for the prevailing road conditions.

Tailgating: Following other vehicles too closely leaves inadequate braking or manoeuvring space.

Erratic Lane Changes: Abruptly changing lanes without signalling, cutting off other drivers, or weaving in and out of traffic.

Abnormal driving can lead to severe traffic conflicts, e.g., crashes, collisions, and other small accidents. Crash frequency and severity are considered two important indicators that directly measure the safety performance of transportation systems. However, crashes are rare events. To address this issue, Surrogate Measures of Safety (SMOS) derived from traffic conflicts have become an increasingly popular solution. SMOS are indirect indicators or metrics used to assess and predict the safety performance of transportation systems, particularly roadways. Unlike direct measures that rely on observed crash data, surrogate safety measures use proxy variables correlated with safety outcomes. These measures are typically collected and analyzed to identify potential safety concerns and prioritize improvement efforts.

Monitoring abnormal driving behaviours in real-time is critical to improving driving safety, raising driver awareness of their driving patterns, and minimizing future road accidents. ML-based approaches have proven to be more effective in abnormal driving behaviour detection than traditional methods because of the ability to learn complex patterns, adapt to changing scenarios, handle large and diverse datasets, detect strange behaviours, and optimize the detection process (Sarker, 2021). These advantages have contributed to the increased adoption of ML in road safety applications and have the potential to enhance the effectiveness of abnormal driving behaviour detection systems significantly. Usually based on labelled data and supervised ML, these methods have shown superior performance to non-machine learning methods. It is also worth noting that most available studies employ only basic motion information, such as vehicle velocity, angle, and coordinates, while neglecting the utilization of complex Safety Measures of Safety (SMOS), such as two-dimensional Time- to-collision.

To bridge the research gaps and address the limitations of existing methods, this thesis aims to explore and develop a novel approach for abnormal driving behaviour detection using a semi-supervised machine learning method. By leveraging the power of semi-supervised machine learning and incorporating SMOS as crucial input features, the thesis intends to enhance the accuracy and effectiveness of abnormal driving behaviour detection.

1.2 Aim of research

Abnormal driving behaviour introduces significant uncertainty to traffic and poses a danger to drivers and the public. Accurately identifying and detecting abnormal driving behaviours are vital in ensuring traffic safety and alerting surrounding vehicles to potential hazards (Jia et al., 2020). Moreover, detecting and removing abnormal driving behaviour from naturalistic driving data are crucial prerequisites for developing human-like driving models for automated vehicles using imitation learning. Abnormal driving behaviour often exhibits complex patterns that traditional rule-based systems may not easily capture. Machine Learning algorithms can learn and understand these intricate patterns through data-driven approaches. They can automatically extract relevant features from raw driving data, enabling the detection of subtle and non-linear relationships that traditional methods might overlook(Sarker, 2021).

The primary objective of this research is to explore, develop, and compare unsupervised and semi-supervised machine learning models for identifying potential abnormal driving behaviour using open-sourced datasets. This study aims to provide meaningful insights into understanding and categorizing human driving behaviour by leveraging these datasets. The results of this research will serve as a foundation for developing human-like driving models for automated vehicles, utilizing empirical data as the basis for training and improving the models' capabilities.

By exploring and developing unsupervised and semi-supervised machine learning models, this research aims to address the limitations of traditional supervised approaches and overcome the scarcity of labelled abnormal driving data. Unsupervised learning algorithms will enable the detection of anomalous patterns and behaviours in the data without the need for pre-existing labels. On the other hand, semi-supervised learning approaches will leverage the limited labelled data available to guide the model in identifying potential abnormal driving behaviour instances. By comparing the performance of these models, this research aims to shed light on the effectiveness of different learning paradigms for detecting abnormal driving behaviour.

Open-sourced datasets ensure the availability and diversity of the data used in this research. These datasets capture real-world driving scenarios, providing a more representative sample of the driving population. By analyzing and categorizing the driving behaviour within these datasets, this study aims to gain insights into the underlying patterns and characteristics of abnormal driving behaviour. This understanding will contribute to developing more accurate and robust models for abnormal driving behaviour detection and pave the way for creating advanced human-like driving models for automated vehicles.

In summary, this research strives to advance the field of abnormal driving behaviour detection by exploring, developing, and comparing unsupervised and semi-supervised machine learning models. Through the analysis of open-sourced datasets, the study aims to provide valuable insights into understanding and categorizing human driving behaviour, ultimately contributing to developing human-like driving models for automated vehicles.

1.3 Societal impact

Through a comprehensive understanding of machine learning techniques for abnormal driving behaviour detection, this research makes valuable contributions to various user groups:

Drivers: The automatic identification of drivers' abnormal driving behaviours can significantly impact promoting safe driving practices. By alerting drivers to their bad driving habits, they can become more aware of their actions and take corrective measures to prevent potential car accidents. This technology is particularly beneficial for novice drivers as it helps them review their driving behaviour and improve their awareness of safe driving practices, enhancing overall road safety.

Vehicle Manufacturers: The detection of abnormal driving behaviour has implications for manufacturers on multiple fronts. Firstly, it can enhance the effectiveness of Driver

Assistance Systems (DAS) technology, specifically designed to improve safety, driving experience, and travel comfort. By integrating abnormal driving behaviour detection into DAS systems, manufacturers can offer advanced features that alert drivers to potential risks and assist them in making safer driving decisions. Additionally, information about driver behaviour can be leveraged to optimize vehicle fuel consumption performance. By detecting and addressing abnormal driving behaviour, manufacturers can develop fuel consumption regulation systems that promote exemplary driving behaviour, leading to optimal fuel efficiency and reduced energy emissions.

Authorities: Abnormal driving behaviour detection results can be valuable for road authorities and enforcement agencies. Road authorities can utilize this information to design or modify infrastructure requirements to enhance safety. For instance, if certain types of abnormal driving behaviour are more prevalent in specific areas or at particular intersections, authorities can make informed decisions regarding traffic signal placement, road signage, or lane configurations to mitigate potential risks. Additionally, enforcing authorities can use the detection results to enforce appropriate measures in cases of aggressive driving behaviour, such as conducting specialized training courses or revoking the driver's license for repeated aggressive driving. This proactive approach can improve overall road safety and discourage dangerous driving practices.

In summary, the insights gained from this research in machine learning for abnormal driving behaviour detection have wide-ranging implications. They benefit individual drivers by promoting self-awareness and safe driving practices, provide opportunities for vehicle manufacturers to enhance Driver Assistance Systems and optimize fuel consumption, and empower authorities to design safer road infrastructure and enforce measures to deter dangerous driving behaviour. Ultimately, these advancements create a safer and more efficient driving environment for all road users.

1.4 Scope of research

This thesis's research scope is focused on analysing and detecting abnormal driving behaviours using available datasets containing detailed vehicle trajectories and motion data. These datasets are essential for studying and understanding abnormal driving behaviours and developing effective detection models.

By utilising these datasets, the researchers can analyse the vehicle trajectories and extract relevant features that capture the characteristics of abnormal driving behaviours. These features include speed, acceleration, lane-changing patterns, distance to other vehicles, and other relevant parameters.

1.5 Thesis outline

The structure of this thesis is organised as follows:

Introduction: A brief overview of the research topic is presented in this chapter, highlighting the importance of detecting abnormal driving behaviours and the potential impact on traffic safety. It also outlines the research objectives and the significance of the study.

Literature Review: The second chapter of the thesis reviews the existing literature on anomaly detection and abnormal driving behaviour detection in transportation. It summarises the research findings on these topics, including the detection features and methods used in previous studies.

Research Gap and Questions: Based on the identified research gaps in the literature review section, the third section formulates two main research questions that align with the research direction of the thesis. These research questions serve as the guiding framework for the subsequent sections.

Data: The fourth chapter focuses on data analysis, where a comprehensive exploration and analysis of the dataset are conducted. This analysis helps identify and characterise abnormal driving behaviours in the dataset.

Methodology: The fifth chapter describes the methodology employed in the study. It introduces the five machine learning models utilised in the research and outlines the experimental processes, including feature ablation research, which investigates the impact of different features on detection performance.

Results, Analysis and Discussion: The sixth chapter presents the results obtained from the experiments using the selected machine learning models. It compares the performance of different models and provides a detailed analysis of the findings.

Conclusions and Recommendations: The seventh chapter discusses the conclusions drawn from the research results. It examines the implications of the findings, discusses the study's limitations, and suggests avenues for future research and improvements.

2 Literature review

Abnormal driving behaviour detection belongs to the category of anomaly detection, so this chapter will first provide an overview of the overall research on anomaly detection. Secondly, previous research on detecting abnormal driving behaviour at the traffic level was summarized, focusing on introducing detection criteria (features) and methods. Research has found that most detection features are basic motion features, such as coordinates, velocity, and acceleration. Only a few studies have mentioned other traffic-related features, such as surrogacy safety measures. In addition, most studies have used simple unsupervised machine learning methods, such as clustering analysis and K-means, without comparing the differences in results between the three types of machine learning: unsupervised, supervised, and semi-supervised.

2.1 Anomaly detection research overview

Abnormal detection is prevalent in traffic driving behaviour. Not only within the transportation domain but across various fields, accurately identifying and detecting anomalies is crucial to ensure safety, optimize performance, and detect unusual patterns or behaviours that deviate from the norm. As industries become automated and connectivity technologies advance, many systems continue to generate massive amounts of data. Anomaly detection, the process of identifying unexpected items or events from data (shown in Figure 1), has become a field of interest for many researchers and practitioners and is now one of the main tasks in data mining and quality assurance (Blázquez-García et al., 2021). Hence, anomaly detection has found diverse applications in a variety of domains, including network intrusion analytics (Ageyev et al., 2021; Jin et al., 2020), medical diagnostics (Charfi & Ansari, 2018; Hajabdollahi et al., 2020), financial fraud protection (Liu, 2022; Singh et al., 2012), manufacturing quality control (Sharifzadeh et al., 2018), marketing and social media analytics (Chae et al., 2012), and more. Here is a nutshell: according to the financial fraud detection methodologies research (Singh et al., 2012), anomaly detection approaches can be categorized in terms of the type of data needed to train the model, i.e., 1) Supervised ML; 2) Unsupervised ML; and 3) Semi-supervised ML to detect financial fraud; As for manufacturing quality control, Sharifzadeh et al., (2018) propose an unsupervised thresholding strategy and a robust supervised abnormality detection strategy. A realistic test scenario with a complex surface geometry is used to assess the performance of the proposed detection algorithm; Convolutional Neural Networks (CNNs) are widely used to analyze abnormalities in medical imaging. Hajabdollahi et al. (2020) propose a bifurcated structure CNN with one branch performing classification and the other performing segmentation.

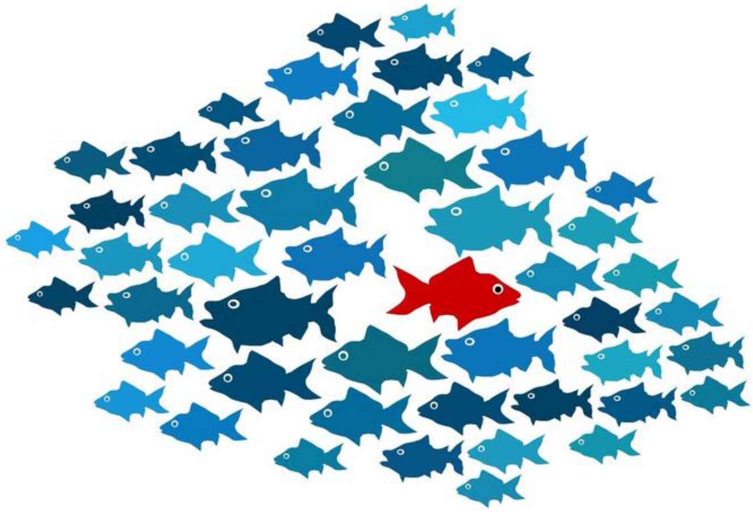

Figure 1 Anomaly detection

2.2 Abnormal driving detection research

This section reviews the existing state-of-the-art technologies for different types of abnormal driving detection and elaborates on the detection features and methods.

2.2.1 Data collection methods

Data used for studying driving style can be collected in several ways, such as driving simulators (Dörr et al., 2014), survey studies (Useche et al., 2019), video (You et al., 2012), and studies based on vehicle motion information (Feng et al., 2018; Jia et al., 2020a; R. et al., 2011). Table 2 provides an overview of the dataset involved in various studies regarding abnormal driving behaviour detection.

Driving simulators

Dörr et al. (2014) developed a driving simulator experiment to collect driver's data and identify the driving style. The navigation system sends a different signal value for dirt tracks, urban streets, rural roads, and motorways. Every road class has a different subsystem for tracing the driving style. So different parameters can be incorporated for every road class. The changeable parameters are the maximum lateral and longitudinal acceleration, the maximum deceleration, the cruising speed, the pedal changing times, and a coefficient of how the driver cuts the curves. For example, When the speed is above the speed limit, the driver is directly classified as sporty. The driver is classified as normal when the speed is below the minimum speed. The driver is classified as comfortable if the speed is between this minimum speed and a specified threshold. Yi et al. (2009) conducted experiments to detect and classify the driving behaviour within the PSAT (Powertrain et al.) environment, a vehicle simulation program developed by Argonne National Laboratory. They consider jerk, which is calculated as the derivative of the acceleration/deceleration or the second derivative of the velocity to measure the driver's aggressiveness. Han et al. (2019) obtained the time-series driving data (i.e.

speed, throttle opening, and acceleration) through a driving simulator to introduce a statistical-based approach to recognize driving behaviour considering driver behaviour uncertainty and develop the Euclidean distance-based decision method to determine the driving style of specific driver behaviour. Daza et al. (2011) designed several experiments in a realistic driving simulator to monitor driver drowsiness based on driver and driving data fusion. Percentage of Eye Closure (PERCLOS), defined as the percentage for 20 seconds for which eyes are at least 80% covered by eyelids, is used with a lateral position and steering wheel angle to monitor the drowsiness in drivers.

Although driving simulators are safe, low-cost, and easy to set up, they only partially represent real-world conditions since it is hard to simulate real-world traffic conditions with all their complexity and variety. Also, drivers might lose spontaneity when they know their driving is monitored.

Survey Studies

The Driver Behaviour Questionnaire (DBQ) is a self-report measure of driving behaviour that has been widely used for over 20 years. In 1990, Reason et al. (1990) introduced the Driver Behaviour Questionnaire (DBQ), which consisted of 50 items describing a variety of errors and violations during driving. Respondents had to indicate how often each aberration occurred during the last year on a scale between 0 (never) to 5 (nearly all the time). Rowe et al. (2015) found that ordinary and aggressive violations were more common in younger people and males in Driver Behaviour Questionnaire. In addition, ordinary violations were a significant independent correlate of crash involvement. Useche et al. (2019) added a new part to the original Driver Behaviour Questionnaire, asking about job-related features and road safety indicators. Same as Driving Simulators, the Survey Study is also not fully representative of real-world conditions, and the question may influence the answer of the respondents.

Video

You et al. (2012) developed the CafeSafe app for Android phones, which fuses information from the front and back cameras and other embedded sensors to detect and alert drivers to dangerous driving conditions in and outside the car. The front camera tracks the driver’s head pose, direction, eyes, and blinking rate to infer drowsiness and distraction. Blinks greater than 500 milliseconds are deemed as indicating micro-sleep. Wei et al. (2013) proposed a drowsy driving detection based on the driver’s physiological signals, such as eye activity measures, the inclination of the driver’s head, sagging posture, response characteristics, decline in gripping force on the steering wheel and lane-keeping characteristics. Algorithms based on video information require installing a camera device/smartphone in the car, and the pre-deployed infrastructure will bring about cost and privacy issues to influence the natural reaction of drivers.

Vehicle motion information

Other studies (Arvin et al., 2021; Jia et al., 2020a; mohammadnazar et al., 2021) used traceable driving information, including vehicle position, speed, and acceleration. The development of connected vehicles (CVs) and location-based services (LBS) provide unprecedented access to information about a driver's location, manoeuvre, speed, and travel time in real-world driving conditions. LBS data are categorized into two groups based on their location acquisition mechanism: 1) data collection via user-end hardware, e.g., smartphones and GPS receivers; 2) data collection via onboard sensors and vehicle-to-infrastructure (V2I) communication. Data from advanced sensors that can transmit and receive Basic Safety Messages (BSMs), which contain information about a vehicle's position, heading, speed, and other information about its state and predicted path, is ideal for tracking traffic conditions and driver behaviour.

High-quality vehicle data is becoming available with the widespread deployment of data collection technologies mentioned above, and the transportation field has entered the era of big data. Furthermore, due to the high dimensionality of the data, traditional statistical methods might not be appropriate in this context. Since the deep learning algorithm has unique adaptability to the time series data, the performance of identifying and predicting data is better than other methods.

Table 1 Dataset overview

Dataset source	Pros	Cons
Driving simulators	Safe, low-cost, and easy to set up	Cannot fully representative of real-world conditions
Survey Studies	Low-cost	Cannot fully representative of real-world conditions
Video	Easy to set up	Cost and privacy issues
Vehicle motion information	High dimensionality of the data	High technical requirements

2.2.2 Abnormal driving behaviour

Various works have been conducted on abnormal driving behaviours (Chen et al., 2015; Hu et al., 2017; Huang et al., 2019; Kim et al., 2016). According to Chen et al., six abnormal driving behaviours are defined (Figure 2), and Kim et al. predefined seven abnormal driving behaviours. The above two definitions of abnormal driving behaviour represent different understandings of driving behaviour between the East and the West (shown in Table 2), led by the United States and South Korea, illustrated in Table 1. The Western driving culture emphasises whether the vehicle's location complies with regulations, while the Eastern driving culture places more emphasis on controlling speed. The different classifications may have a lot to do with the East and West population density. For example, eastern countries such as China, Japan, and South Korea have higher population density and higher numbers of vehicles.

Due to the previous review, this thesis's definition of abnormal driving behaviours will combine location and speed. The abnormal driving behaviours focused on are (1) Sudden start, (2) Emergency braking, (3) Rapid Lane changing, (4) Close Lane changing, and (5) Oppositional conflict.

Table 2 Different definition of Abnormal Driving Behaviour by Chen et al., 2015; Kim et al., 2016

U.S. NHTSA	Korea MOLIT
Weaving	Sudden start
Swerving	Speeding
Sideslipping	Long-standing speeding
Fast U-turn	Sudden braking
Turning with a wide radius	Sudden overtaking
Sudden braking.	Sudden changing lanes
	Sudden turning

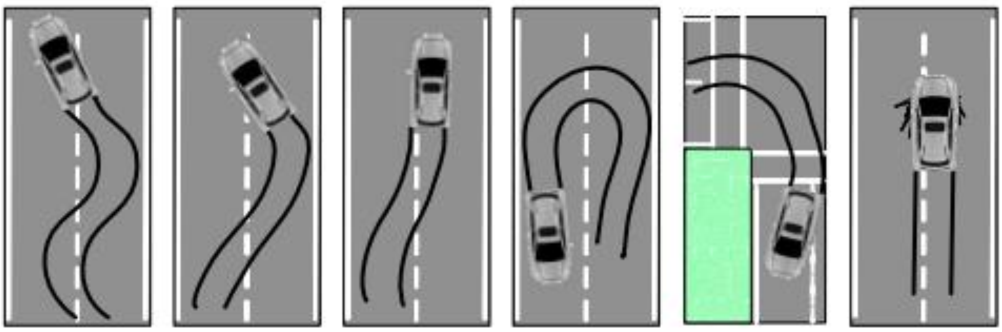

Figure 2 Abnormal driving behaviour(Chen et al., 2015)

2.2.3 Traditional featuress used in ML-based methods

With the advancement of machine learning (ML) and artificial intelligence techniques, ML-based abnormal detection methods have rapidly risen and delivered super performance. To make ML models work, features must be provided and fed into the model. Various traditional indicators have been adopted as input features.

Plank et al. (2015) collect lateral vehicle position, vehicle steering angle and speed-related data from the car simulator and use the SVM to classify the two driving states as normal and drunken.

Lim & Yang (2016) use vehicular data considered, including velocity, lateral and longitudinal acceleration, steering angle, gas pedal angle, etc., for estimating drowsiness, high workload, and drivers' cognitive and visual distraction using the convolutional neural network model.

In another study performed in 2010 (Dai et al., 2010), acceleration and orientation data collected from smartphone sensors were used to differentiate between aggressive and non-aggressive driving. In 2012, images extracted from vehicle cameras were used for detecting abnormal driving by identifying dangerous events like a sudden lane change, brake, sudden acceleration, or sudden deceleration.

Later, the LSTM-CNN-based prediction model was used for abnormal driving detection from vehicular data, such as GPS, throttle position, acceleration etc. (Jia et al., 2020b).

Different types of driving behaviour abnormality have been detected by Dhar et al. (2014), such as the lane position maintenance problem, which includes drifting, swerving, abrupt U-turn, etc., and speed control problems, including sudden acceleration, sudden deceleration, braking, abrupt stopping, etc.

Table 3 provides a comprehensive overview and summary of the various features described in the literature review.

Table 3 Traditional features overview

Problem Considered	Author	Indicators
Drunken Driving Detection	Planek et al., (2015)	Vehicle lateral position/Steering angle/Velocity
Distracted Driving Detection	Lim & Yang, (2016)	Velocity/Lateral and longitudinal acceleration/Steering angle/Gas pedal angle
Aggressive Driving Detection	Dai et al., (2010)	Acceleration/Orientation
Abnormal Driving Detection	Jia et al., (2020)	GPS/Throttle position/ Acceleration
Driver State Recognition	Dhar et al., (2014)	Acceleration/Deceleration/Braking

2.2.4 Surrogate measures of safety used in ML-based methods

In addition to the coordinates, velocity, vehicle angles and other traditional features, surrogate measures of safety are also used in ML-based methods to detect abnormal driving behaviour.

Rong Chen & Rini Sherony (2016) used Time to Collision (TTC) to detect abnormal driving behaviour and verify the effectiveness of Forward Collision Warning (FCW) in their study. Time to collision (TTC) is one of the most widely used SSMs introduced by Hayward (1972) and has been used in different studies to evaluate the risk of a rear-end collision. If two vehicles continue along the same path at their present speed, the collision time is calculated as follows:

$$TTC = \begin{cases} \dfrac{s_0 - l}{v - v_0}, & v > v_0 \\ \infty, & otherwise \end{cases}$$

where s_0 is the space headway between the following and leading vehicles, l is the length of the leading vehicle, v_0 and v are the initial velocity of the leading and following vehicles, respectively. If the TTC value is less than a threshold, the car-following scenario is considered to be unsafe.

2.2.5 Machine learning model

Recently, there have been some studies on abnormal driving behaviour using clustering and shallow learning algorithm, which can only classify drivers' driving styles without identifying specific types of abnormal behaviour (Bejani & Ghatee, 2018; Suzdaleva & Nagy, 2018). Deep learning has recently received significant attention in academic and industrial circles as a new state-of-the-art machine learning approach. It uses a multiple-layer architecture and is pre-trained for extracting inherent features from vast amounts of unlabeled data. The available methods used in previous abnormal driving behaviour research can be grouped into supervised and unsupervised methods.

Supervised

For supervised machine learning to succeed, input and output data must be labelled during the training phase. Before the training and testing phases of the model, a data scientist labels this training data. The model can classify and predict new datasets by learning the relationship between inputs and outputs. It is called supervised machine learning because at least part of this approach requires human oversight. The vast majority of available data is unlabelled, raw data. Labelled data is generally required to be ready for supervised learning accurately. Naturally, this process can be resource-intensive, as large arrays of accurately labelled training data are needed. Figure 3 is a schematic diagram of the supervised model operation.

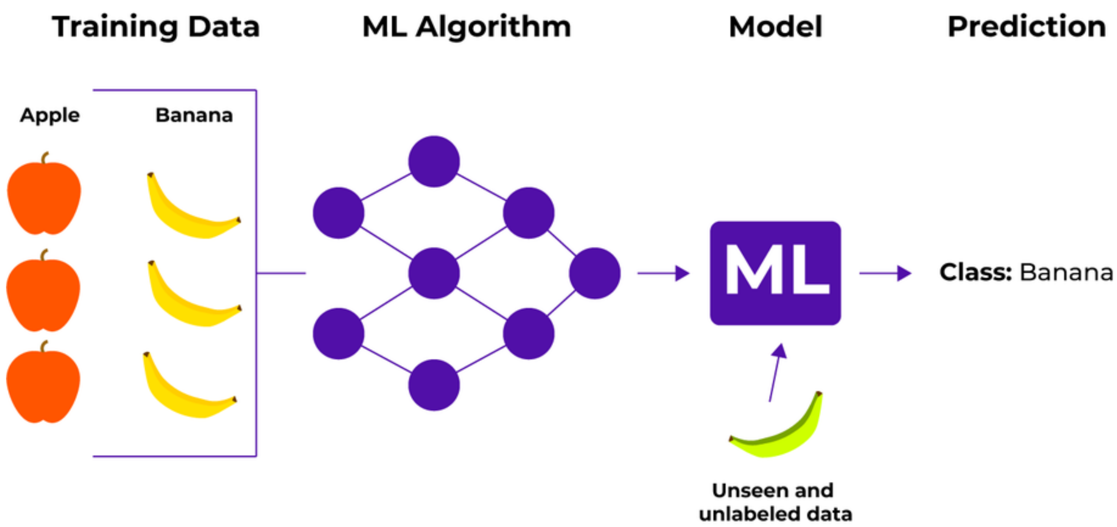

Figure 3 Example of how supervised learning works
(https://neurospace.io/blog/2020/08/what-is-supervised-learning)

Arvin et al. (2021) applied a 1D-Convolutional Neural Network (1D-CNN), Long Short-Term Memory (LSTM), and 1DCNN-LSTM to capture the local dependency and volatility in time-series data. Jia et al. (2020) built a long short-term memory network and convolutional neural network (LSTM-CNN) model based on the advantages of LSTM in processing time series data and CNN in processing matrix data. The extreme acceleration and deceleration points are detected through the statistical analysis of actual vehicle driving data, and the driving behaviour recognition data set is established. By using the data set to train the model, the LSTM-CNN can achieve a better result. A lightweight 1D Convolutional Neural Network with high efficiency and low computational complexity was suggested to classify the driver behaviour in the study of Shahverdy et al. (2021). Moving at high speed, braking, rapidly changing the speed, and quick steering are the activities that the study concerned. Ryan et al. (2021) simulated an end-to-end model of Autonomous vehicles (AV) by using Convolutional Neural Networks (CNN) to compare human and AV driving behaviour.

Unsupervised

As the name suggests, unsupervised machine learning is more hands-off than supervised machine learning. Unsupervised machine learning is training models on raw and unlabeled training data. It's also often an approach used in the early exploratory phase to better understand the datasets. Figure 4 is an example of an unsupervised model.

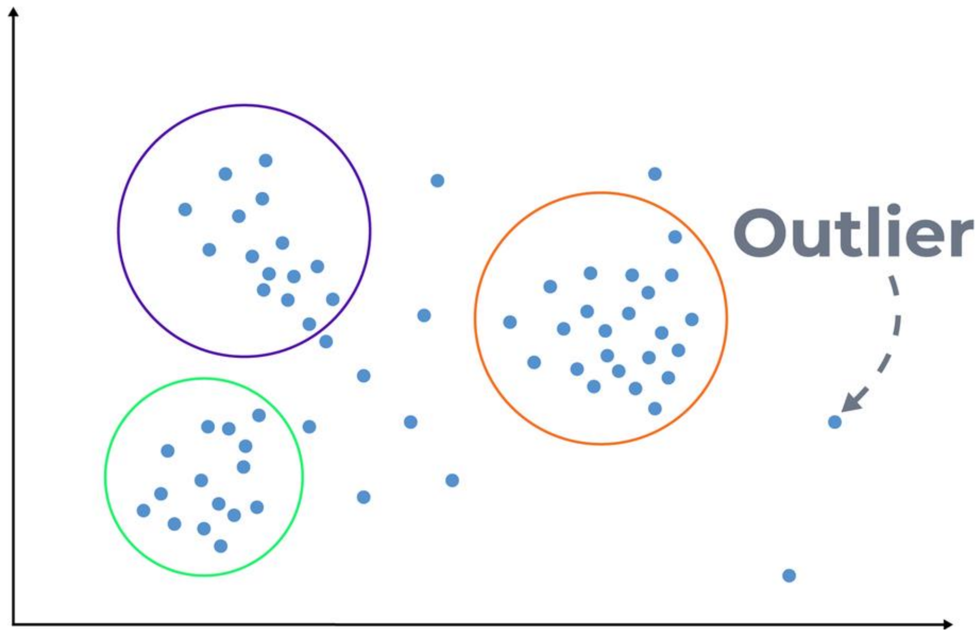

Figure 4 Example of Unsupervised Learning
(https://neurospace.io/blog/2020/10/what-is-unsupervised-learning/)

Mohammadnazar et al. (2021) developed a framework to quantify instantaneous driving behaviour and classify driving styles in different spatial contexts using unsupervised machine learning methods. To quantify driving style, the concept of temporal driving volatility, as a surrogate safety measure of unsafe driving behaviour, was utilized and applied in this study. K-means and K-medoid methods are applied for grouping drivers in aggressive, normal, and calm clusters. Feng et al. (2018) proposed a novel technique to robustly classify driving style using the Support Vector Clustering approach, which attempts to differentiate the variations in individuals' driving patterns and provides an objective driver classification. Four input signals (vehicle speed, engine speed, pedal position, and headway distance) and four typical statistical features (mean, standard deviation, maximum and minimum values) were identified as the feature parameters.

Semi-supervised

Semi-supervised machine learning is an approach that falls between supervised and unsupervised learning. In semi-supervised learning, the training data consists of labelled examples (data points with assigned labels) and unlabeled examples (data points without assigned labels). Figure 5 shows an example of one of the semi-supervised models.

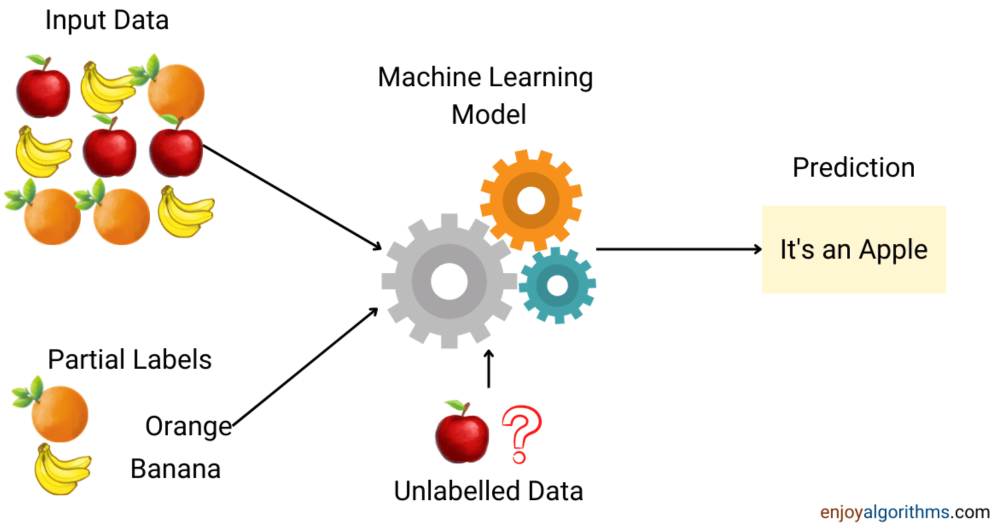

Figure 5 Example of Semi-supervised Learning (https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning)

Semi-supervised models have been widely used in anomaly detection, especially for network safety: Lin & Chiang (2022) introduces SNetAD, a novel semi-supervised approach for anomaly detection in large network logs. Bilal Hussain et al. addresses the challenges posed by increasing network complexity and the underutilization of big data in mobile networks by proposing a semi-supervised statistical-based anomaly detection technique.

However, it is rarely applied in the field of transportation. Oikawa et al. investigate the effectiveness of an Online Sequential Extreme Learning Machine (OS-ELM) for detecting anomaly driving behaviour using sensor data, comparing its performance with Hidden Markov Model (HMM) and Long Short-Term Memory (LSTM) methods, and demonstrates that the OS-ELM-based detector achieves comparable or better accuracy in anomaly detection with faster sequential learning speed. An Online Sequential Extreme Learning Machine (OS-ELM) is a semi-supervised model.

Table 4 summarizes the machine learning models and their accuracy in different literature.

Table 4 Overview of Machine Learning Applied for Abnormal Driving Detection

Problem Considered	**Author**	**Algorithms used**	**Classification Accuracy**
Driver State Estimation	Eren et al., (2012)	Optimal path detection algorithm, Bayesian	93.3%

		classifier	
Aggressive Driving Detection	Karaduman et al., (2013)	SVM, k-means clustering	Unspecified
Aggressive Driving Detection	Van Ly et al., (2013)	SVM, k-means clustering	Around 60%
Aggressive Driving Detection	Castignani et al., (2015)	PCA, SVM, k-means clustering	90%
Driving Behavior Profiling	Ferreira et al., (2017)	Artificial Neural Network, Random Forest	Unspecified
Abnormal Driving Detection	Jia et al., (2020)	Statistical analysis, LSTM-CNN	95%
Abnormal Driving Detection	Hiroki Oikawa et al., (2020)	Online Sequential Extreme Learning Machine (OS-ELM)	

2.3 Conclusion

This chapter comprehensively overviews data collection methods, including video recordings, surveys, and driving simulators. It highlights the advantages and limitations of each method in capturing driving behaviour data for anomaly detection purposes. Additionally, the chapter summarises the classifications of abnormal driving behaviours, shedding light on the various types of driving anomalies identified and studied. Furthermore, it examines the features employed in machine learning-based anomaly detection approaches, emphasising basic motion information while often neglecting the potential of other significant factors, such as Surrogate Measures of Safety (SMOS). Lastly, the chapter explores the application of different machine learning models in anomaly detection for driving behaviour, noting the widespread use of unsupervised and supervised models while highlighting the comparatively limited utilisation of semi-supervised models in this domain.

3 Research Gap and Research Questions

3.1 Research gaps

From the literature review chapter, the research gaps can be summarized from two aspects: detection methods and detection features:

Most anomaly detection methods use shallow machine learning, such as clustering and K-means, with only a small portion using deep learning methods. In addition, unsupervised and supervised methods are often used in different types of machine learning, while semi-supervised methods are often overlooked. At the same time, more is needed to compare the results of different types of machine learning.

In traditional methods (clustering, K-means, etc., mentioned in the previous point), the input features include basic motion features but ignore SMOS. So, in this thesis, whether SMOS also help with machine learning methods needs to be explored.

3.2 Research Questions

Following the identified research gaps, the two main research questions of this study are raised:

How can abnormal driving behaviour be accurately and effectively detected?

To answer it, three sub-questions are developed:

1.1 What are abnormal driving behaviours? What types of abnormal driving behaviours are present in this dataset?

1.2 What are the results of machine learning methods for detecting abnormal driving behaviour?

1.3 What are the differences in the performance of unsupervised, supervised, and semi-supervised machine learning?

1.4 What are the important features regarding abnormal behaviour detection?

1.5 Will SMOS help to improve the model performance?

1.6 What are the changes in the results of different machine learning models after the introduction of SMOS?

1.7 Can the model still detect abnormal driving behaviour without using SMOS?

Sub-question 1.1 aims to identify and define abnormal driving behaviours and explore the specific types observed in the dataset used in the study. It seeks to understand the range of abnormal driving behaviours that must be detected.

Sub-question 1.2 evaluates the performance and effectiveness of different machine learning methods in detecting abnormal driving behaviour.

Sub-question 1.3 compares and contrasts the performance of different machine learning approaches in detecting abnormal driving behaviour, specifically unsupervised, supervised, and semi-supervised methods. It investigates the strengths and limitations of each approach and examines how they differ in terms of accuracy, efficiency, and the need for labelled training data.

Sub-question 1.4 and 1.5 explores the potential benefits of utilizing SMOS (Surrogate Measures of Safety) as a feature in the abnormal behaviour detection model. It investigates whether including SMOS as a feature enhances the model's performance in accurately identifying abnormal driving behaviour.

Sub-question 1.6 and 1.7 examines the impact of incorporating SMOS as a feature on the outcomes of various machine learning models. It compares the performance of different models (e.g., Random Forest, Isolation Forest and HELM) with and without the inclusion of SMOS, assessing any improvements or changes in accuracy or other performance metrics.

4 Data

In this chapter, sub-question 1.1 will be answered. First, a comprehensive dataset description is illustrated, and then abnormal driving behaviours at different locations are identified.

4.1 Dataset requirements

The selection of a good data set that suits the purpose of research requires the definition of criteria. In general, look for data sets that meet the following conditions:

Useable: The dataset should be available as open-source data.

Contain the needed elements: The dataset should have basic motion information such as velocity, acceleration, location (coordinate), timestamp, etc.

Have two dimensions: Only having longitudinal or latitudinal data makes detecting abnormal behaviour challenging and sometimes impossible. Thus, a 2D dataset is necessary.

Have huge metadata: Since there may be invalid data, the dataset size should be large enough. Also, data need to be separated into training, validating, and testing subsets.

Abnormal driving behaviour: The dataset should contain at least some abnormal driving behaviour for further research.

4.2 Dataset description

Three candidate datasets are already studied by previous research, so it is necessary to find abnormal driving behaviour from these datasets:

One of the candidate datasets is The Safety Pilot Model Deployment Data (SPMD), which contains the corresponding vehicle data by the US Department of Transportation Security participating in Ann Arbour, Michigan. Containing information from 2836 vehicles equipped with V2V technology and 30 roadside equipment (RSE) covering more than 73 lane miles on public streets, the SPMD is one of the most extensive real-world data collection programs ever undertaken in the field (Bezzina & Sayer, 2014).

Another dataset candidate is the second Strategic Highway Research Program (SHRP2), which contains more than 4 petabytes of information, known as the most comprehensive driving study. The data collection was performed from 2010 to 2013 and contained high-quality, high-resolution data from six states. The data contains information about 3500 drivers with more than 50 million miles travelled (Hankey

Miguel et al., 2016).

Thirdly, CitySim is a video-based trajectory dataset generated from drone recordings focusing on traffic safety in the United States. It contains vehicle trajectory data extracted from 12 different drone videos recorded over 1140 minutes. This textbook covers six types of road geometry: freeway segments, weaving segments, merge/diverge segments on expressways, signalized intersections, stop-controlled intersections, and intersections without signs or signals (Zheng et al., 2022).

Based on the above conditions, this thesis selects CitySim as the dataset for research. Because there is a certain amount of abnormal driving behaviour in this dataset, the collected information is easier to calculate the required features for this thesis.

The CitySim dataset trajectories are provided as Comma Separated Value (CSV) files. Each row represents a waypoint that belongs to a vehicle trajectory in a single frame. Each waypoint contains the position information of seven vehicle key points: centre point, head, tail, and four bounding box vertices, as depicted in Figure 6. The dataset used in this study provides position information in multiple formats, including pixels, feet, and GPS coordinates. Additionally, it includes data on speed, heading (measured concerning both the global north and the image X-axis), and the vehicle lane number. It is important to note that the accuracy of the dataset is within a range of approximately 10 centimetres, indicating a high level of precision in the recorded measurements.

However, only the above features cannot meet the data support the model requires. Further calculations in this thesis add additional features such as acceleration, lateral acceleration, the distance between two vehicles, and Time-to-collision (TTC) to the original dataset.

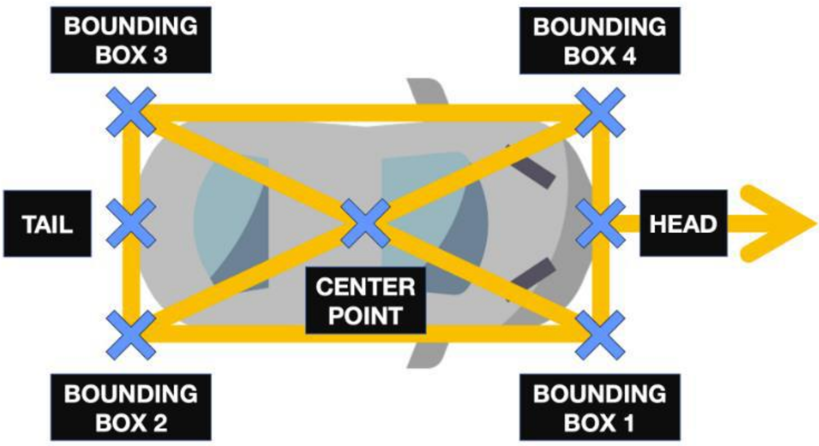

Figure 6 Vehicle bounding box feature description

Table 5 provides an example of the raw data available in the dataset. It includes the following features and their corresponding values:

- frameNum: The frame number is 0.
- carId: The car identifier is 582.
- carCenterX: The x-coordinate of the car's center position is 462.4 feet.
- carCenterY: The y-coordinate of the car's center position is 184.8 feet.
- headX: The x-coordinate of the car's head position is 469.6 feet.
- headY: The y-coordinate of the car's head position is 184.8 feet.
- tailX: The x-coordinate of the car's tail position is 455.3 feet.
- tailY: The y-coordinate of the car's tail position is 184.8 feet.
- Speed: The car's speed at this data point is 39.5 miles per hour.
- Heading: The car's heading direction is 180.7 degrees.
- laneId: The car is located in lane number 10.

Table 5 Example of dataset

Feautres	**Unit**	**Value**
frameNum	/	0
carId	/	582
carCenterX	Ft	462.4
carCenterY	Ft	184.8
headX	Ft	469.6
headY	Ft	184.8
tailX	Ft	455.3
tailY	Ft	184.8
Speed	mph	39.5
Heading	degree	180.7
laneId	/	10

4.2 Abnormal driving behaviour in the dataset

According to the classification and definition of abnormal driving behaviour in the literature review (seen in 2.2.2), the abnormal driving behaviour in this dataset is illustrated as follows in this section. Each abnormal driving behaviour corresponds to one or two indicators for judgment, which need to be measured or calculated in different locations. Figure 7 shows the relationships among behaviour, indicators and location.

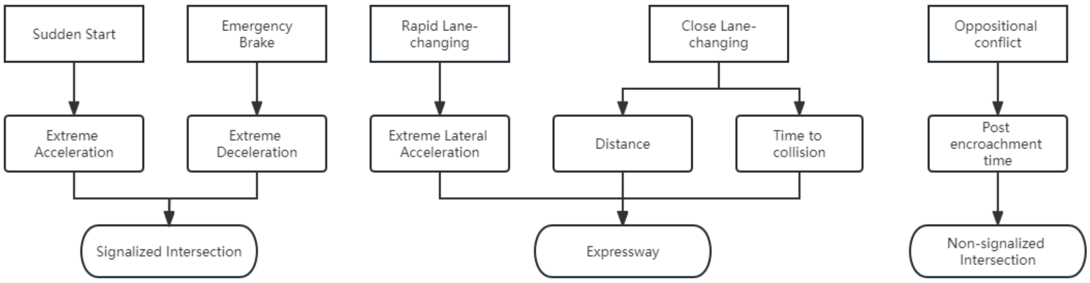

Figure 7 Relationship between behaviour, indicators and location (drawn by *Zhang lanxin*)

4.2.1 Signalized Intersection

The first location selected is a signalized intersection near Alafaya University in Florida, as seen in Figure 8. Due to the existence of traffic lights, sudden start and emergency braking will probably occur at this place.

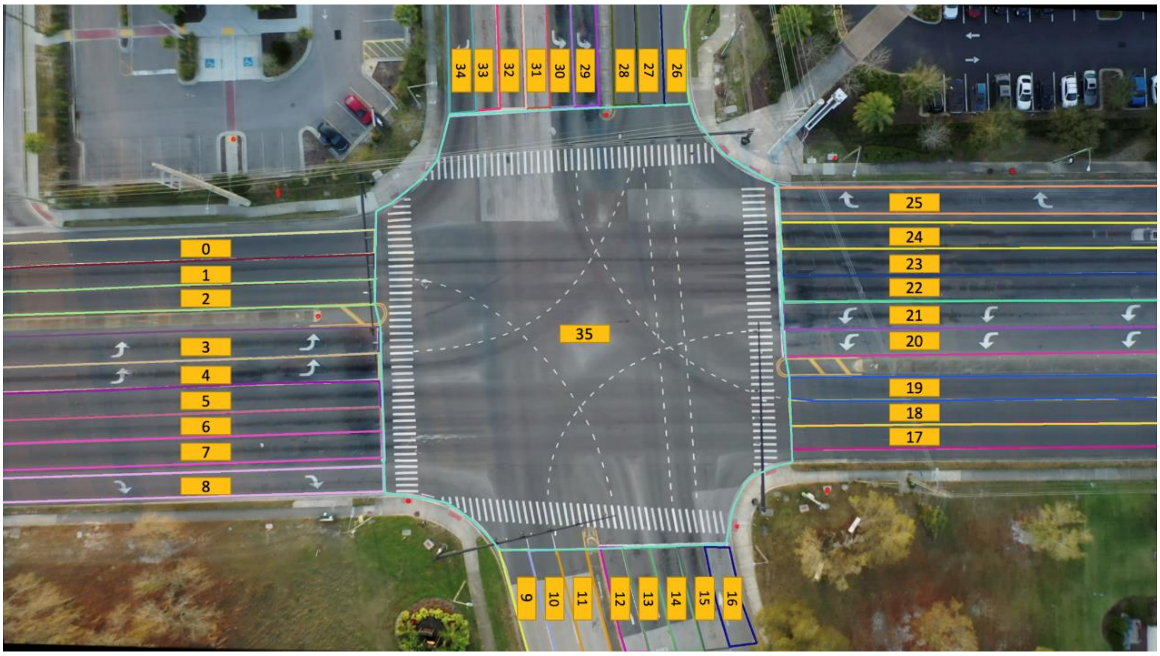

Figure 8 Alafaya University Signalized Intersection

The data collection information is shown in Table 6, and there are, in total, 414,976 data instances involved in the analysis for this selected signalized intersections.

Table 6 Data collection information at Alafaya

Location	**Location Type**	**Drone Height (m)**	**FPS**	**Recording Resolution**	**Recording Length (min)**
University Alafaya	Signalized Intersection	120	30	3840 x 2160	120

For all vehicle trip data, much acceleration data correspond to each speed, which means that the acceleration data at each speed can be statistically analysed, and the extreme acceleration and deceleration points at each speed can be calculated. The acceleration changes sharply when the driver performs abnormal operations (such as sudden braking and accelerating). Figure 9 shows the specific manifestation of rapid acceleration and emergency braking in real life. Therefore, the abnormal driving behaviour of the driver can be segmented by detecting the extreme points of acceleration. A certain proportion of extreme acceleration points can be selected by statistically analysing all the acceleration points at the same speed in all trips. At a given speed, most acceleration points are caused by normal driving operations, while abnormal driving behaviours cause a small number. Therefore, a proportion should be set to separate extreme acceleration points from normal ones. If the ratio is too high, then most of the acceleration points will be considered normal, resulting in some abnormal behaviours that cannot be detected. On the contrary, many normal operations will be recognised as abnormal. Therefore, through repeated experiments and related research (Jia et al., 2020; Wang et al., 2015), it is reasonable to set the proportion to 16% since, for a bell-shaped normal speed distribution, 68% of the mass will be within one standard deviation, see in Figure 10.

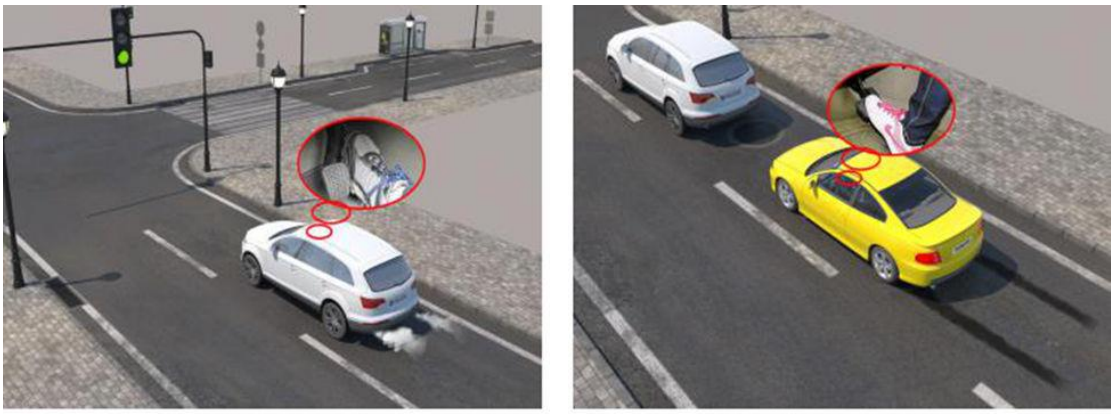

Figure 9 Rapid acceleration and Emergency braking (Jia et al., 2020)

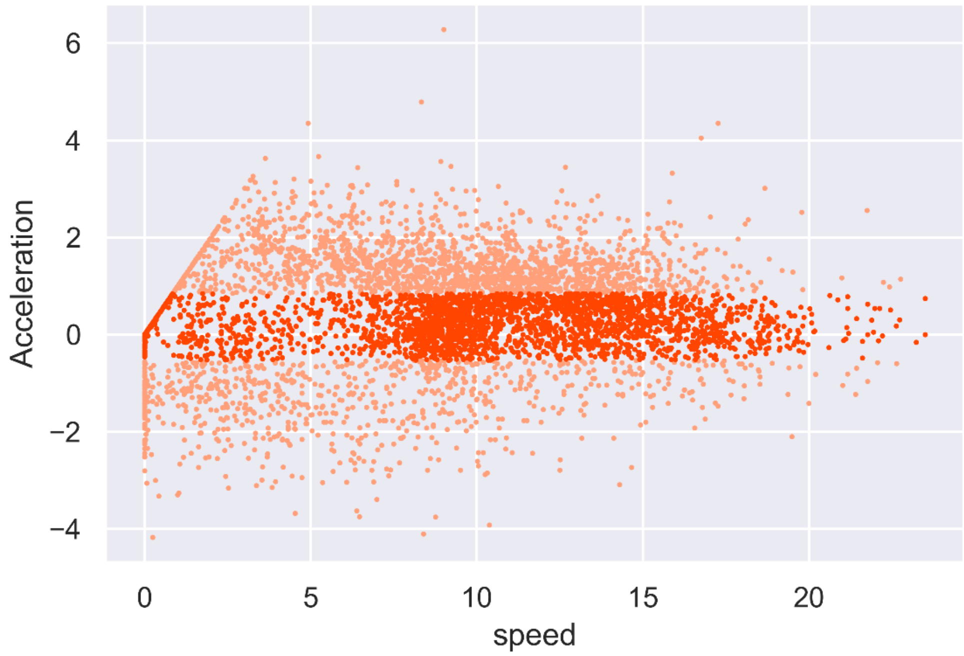

Figure 10 Extreme acceleration and deceleration points distribution at different speeds (red=normal/orange=abnormal)

4.2.2 Expressway

Expressway A is selected as another analysis case for abnormal lane-changing behaviours. The weaving segment of Expressway A (seen in Figure 11) exhibits many critical safety events such as cut-ins, merges, and other lane-changing behaviours. In this section, rapid lane-changing and close lane-changing are introduced.

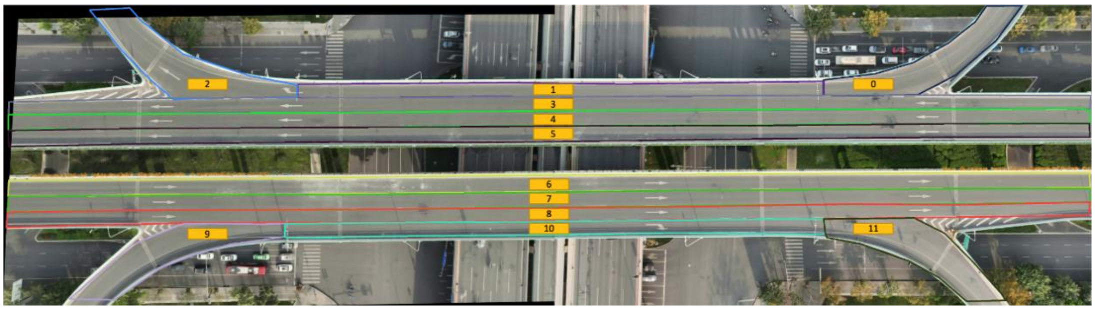
Figure 11 Expressway A

The data collection information is shown in Table 7; there are 260,690 data at Expressway A.

Table 7 Data collection information at Expressway A

Location	Location Type	Drone Height (m)	FPS	Recording Resolution	Recording Length (min)

Expressway A	Weaving Segment	120	30	5120 x 2880	120

A lane-changing vehicle which cuts in from a source to a target lane may cause a significant conflict with the following vehicle on the target lane. It is one of the driving behaviours that cause rear-end crashes. Since the studying areas of Expressway A are within weaving segments, it is unsurprising that there are lane change behaviours here.

Table 8 Lane changing behaviour at Expressway A

ExpresswayA	**No lane changing**	**Once**	**Twice**	**Three times**	**Four times**	**Total**
Lane-changing	325	97	118	47	5	592

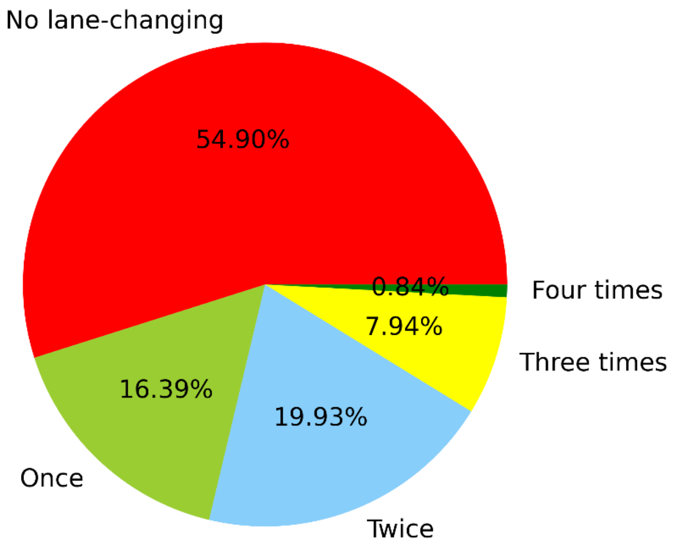

Figure 12 Proportional distribution of lane changing times

Based on the given proportions, we can analyze the distribution of lane-changing behaviours in the dataset:

- **No lane changing**: Approximately 54.9% of the instances in the dataset did not involve any lane changing, which indicates that most observed driving behaviours did not include lane changes.
- **Once**: Around 16.4% of the instances involved a single-lane change, suggesting many driving instances included a single-lane change during the observed period.
- **Twice**: Approximately 19.9% of the instances involved two lane changes, indicating a relatively higher frequency of instances where two lane changes occurred during the observed period.
- **Three times**: Roughly 7.9% of the instances involved three-lane changes, which suggests a lower frequency of instances with three-lane changes.
- **Four times**: Only around 0.8% of the instances involved four-lane changes, which indicates a very low occurrence of instances with four-lane changes.

Figure 12 shows the proportional distribution of lane-changing times using a pip chart. The analysis reveals that most instances in the dataset did not involve lane changes, followed by instances with a single lane change. Instances with multiple lane changes (two or three) were less frequent, and instances with four-lane changes were the least common.

Rapid Lane Changing

For the vehicles with lane-changing behaviour, the lateral acceleration is analysed to see whether the vehicles are rapid lane-changing. Figure 13 shows the specific manifestation of rapid lane-changing behaviour in real life.

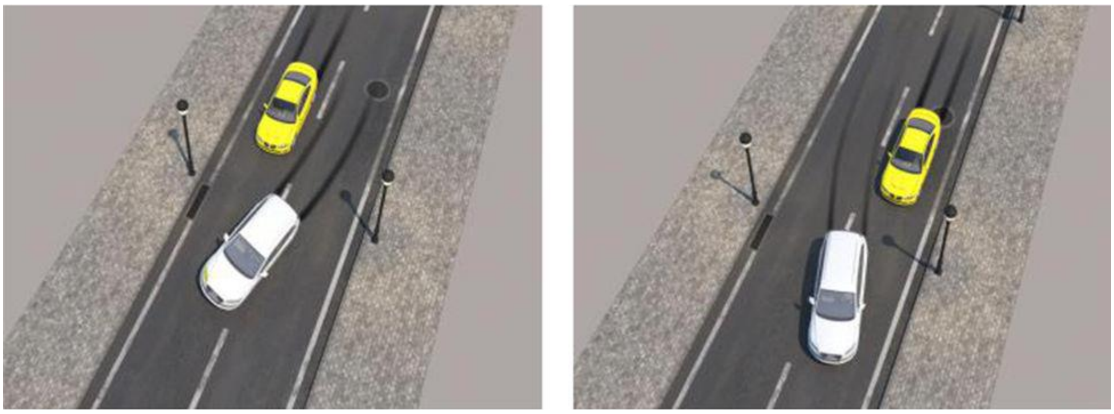

Figure 13 Rapid lane changing (Jia et al., 2020)

The following Table 9 and Figure 14 are the analysis of lateral acceleration for the vehicles that have lane-changing behaviour.

Table 9 Lateral acceleration

Expressway A	Mean	Std	Min	Max
Lateral Acceleration(m/s^2)	0.00	1.30	-6.59	6.42

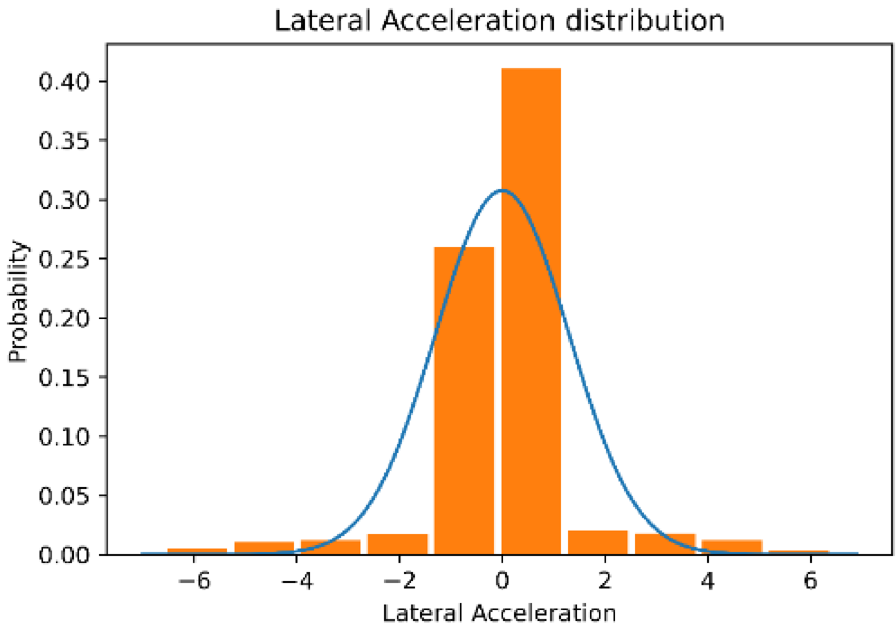

Figure 14 Lateral Acceleration Distribution

Most vehicles with lane change behaviour have an acceleration of about 0 m/s ², which means they change lanes at a constant speed. However, the acceleration of some vehicles is an outlier seen in Figure 15. According to the normal distribution, a value greater than 1.3m/s ² and less than -1.3m/s ² will be the filter condition for outliers.

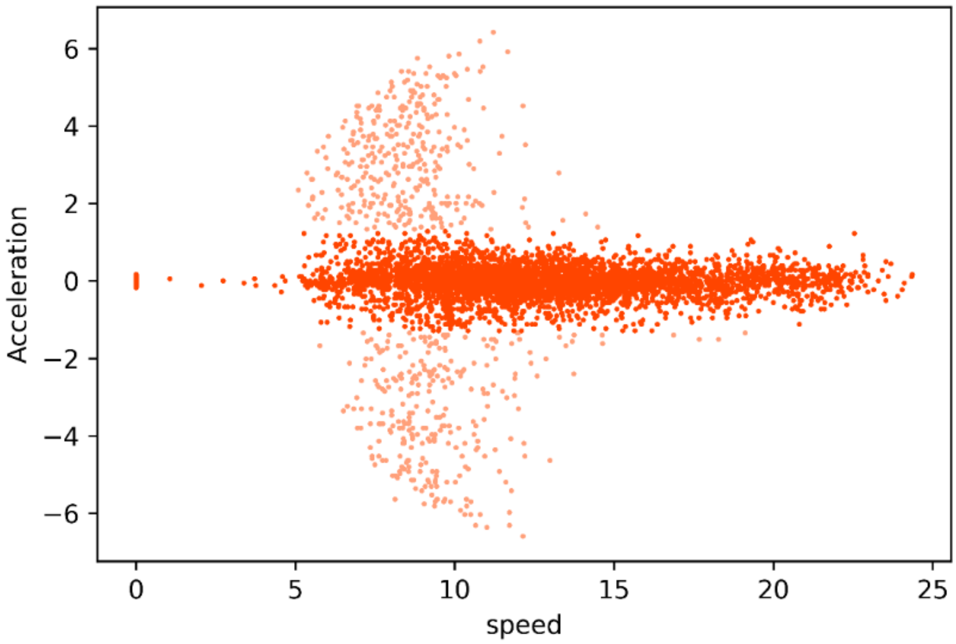

Figure 15 Extreme lateral acceleration and deceleration points distribution at different speeds (red=normal/orange=abnormal)

Close lane changing

Hazards can also arise when a vehicle changes lanes and gets too close to other vehicles. In this situation, abnormal driving behaviour is detected by the distance between two vehicles with lane-changing behaviour. Figure 16 shows the specific manifestation of close lane-changing behaviour in real life.

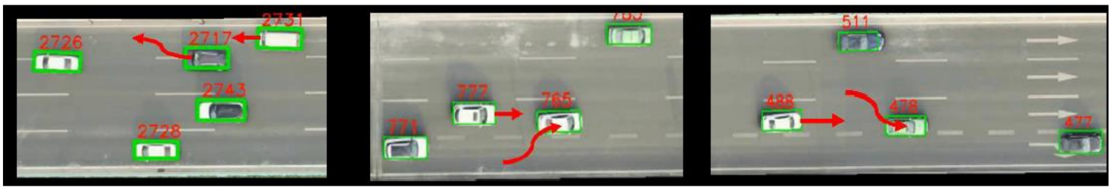

Figure 16 Close Lane changing (Zheng et al., 2022)

Figure 17 shows the distance between all vehicles on other roads and the lane-changing vehicles when any vehicle on the road changes lanes.

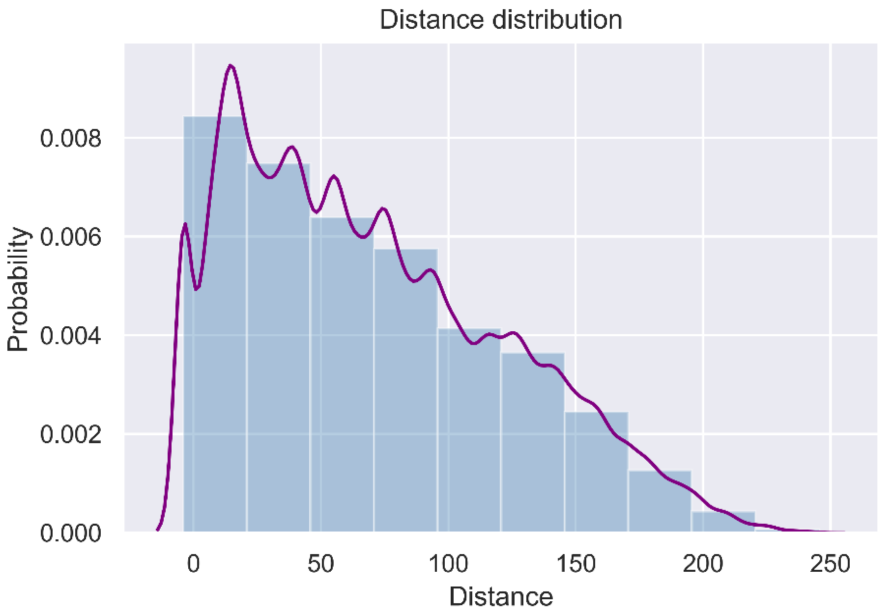

Figure 17 Distance distribution in Expressway A

In this thesis, when the distance between two cars during lane-changing is less than 0.5 meters, it is considered severe abnormal driving behaviour. In contrast, when the distance is less than 1.0 meters but greater than 0.5 meters, it is considered weak abnormal driving behaviour, as seen in Figure 16.

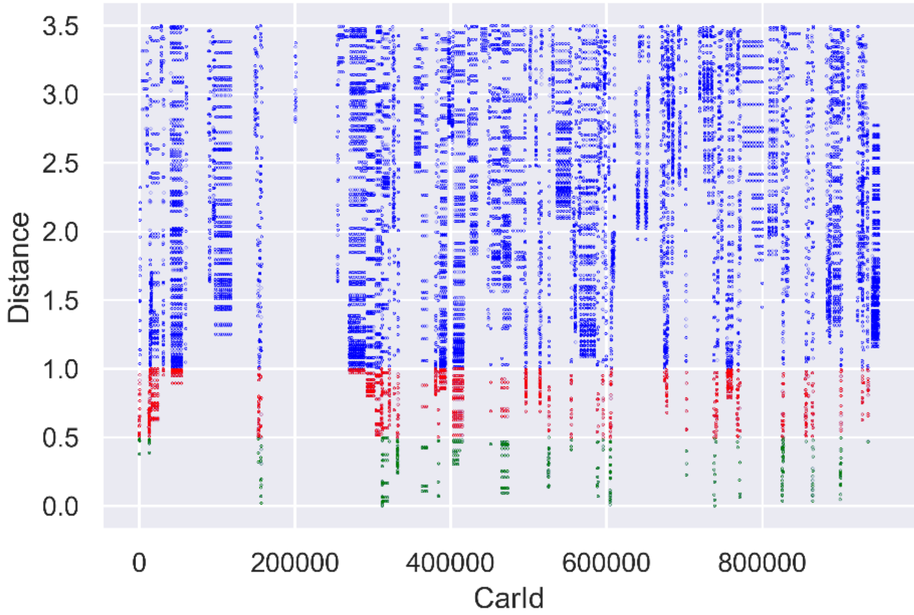

Figure 18 Extreme distance points distribution at different CarID (green<0.5m/0.5 ≤red≤1.0m/blue>1.0m)

4.2.3 Non-signalized Intersection

One observable measure that allows for consistency between observers and locations is post-encroachment time (PET). The McCulloch Seminole intersection (see in Figure 19) is non-signalized, and the yellow areas in the figure below are the range where conflicts are likely to occur at this intersection.

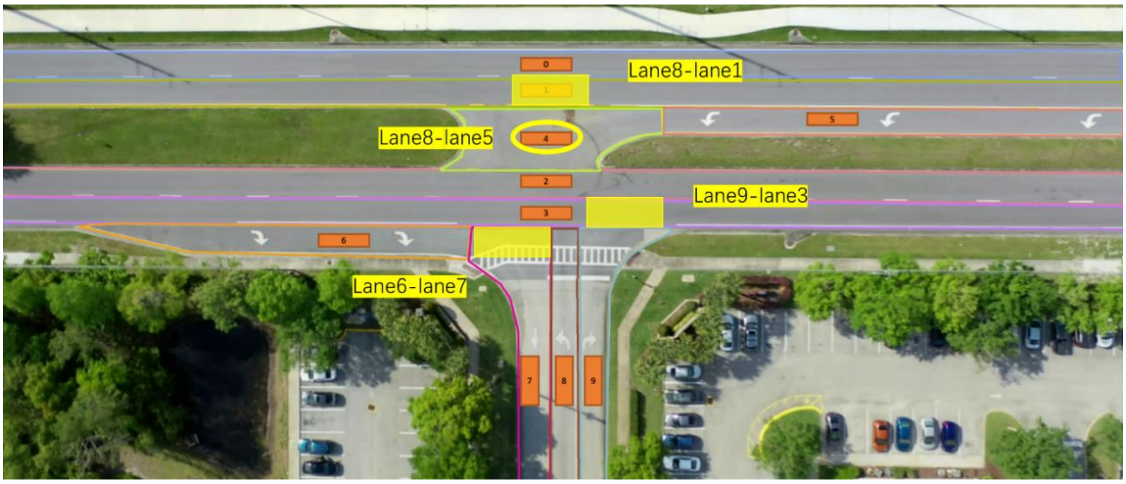

Figure 19 McCulloch Seminole Non-Signalized Intersection

The data collection information is shown in Table 10, and there are 335,660 data at Non-signalized Intersection.

Table 10 Data collection information at University McCulloch

Location	Location Type	Drone Height (m)	FPS	Recording Resolution	Recording Length (min)
University McCulloch	Control-Free Intersection	120	30	3840 x 2160	60

However, despite the abundance of data available at non-signalized intersections, it is worth noting that only two conflicts occurred within the four areas above, as seen in Table 11.

Table 11 Post Encroachment Time

CarID	Conflict Zone	Post Encroachment Time (PET)
233-267	Lane9-Lane3	1.6s
105-111	Lane9-Lane3	1.87s

4.3 Conclusion

This chapter provides an overview of the dataset and a summary of different abnormal driving behaviours observed at various locations. The specific characteristics of abnormal driving behaviours at signalized intersections, unsignalized intersections, and expressways are described. The identified abnormal driving behaviours in the dataset include rapid acceleration, emergency brake, close lane-changing, and rapid lane-changing.

After the analysis above, the dataset used in this thesis contains three types of abnormal driving behaviours, as shown in Table 12.

Table 12 Abnormal driving behaviour summary in this thesis

<table>
<tr><td>Emergency brake/Sudden start</td><td colspan="2">Normal</td><td colspan="2">Abnormal</td></tr>
<tr><td>Rapid lane changing</td><td colspan="2">Normal</td><td colspan="2">Abnormal</td></tr>
<tr><td>Close lane changing</td><td>Normal</td><td colspan="2">Weak abnormal</td><td>Severe abnormal</td></tr>
</table>

5 Methodology

In the previous chapter, the dataset was labelled as either normal or abnormal, which served as the annotation task for the subsequent machine learning analysis (e.g. unsupervised and semi-supervised). Labelling was conducted to assign the appropriate classification to each data instance, providing the necessary groundwork for the machine learning algorithms used in the following stages. This chapter describes the research methodology and experiment setup in detail, including an introduction to the model, experimental design and experimental procedure.

5.1 ML models

5.1.1 Supervised ML

Supervised anomaly detection algorithms aim to incorporate application-specific knowledge into the detection process. This thesis will introduce two supervised machine learning models to detect abnormal driving behaviour.

XGBoost

The first supervised machine learning model is a scalable end-to-end tree boosting system called XGBoost, used widely by data scientists to achieve state-of-the-art results on many machine learning challenges. XGBoost was mainly designed for speed and performance using gradient-boosted decision trees. It represents a way for machine boosting, or in other words, applying to boost to machines, initially done by Chen & Guestrin (2016) and further taken up by many developers.

The XGBoost algorithm optimises an objective function by iteratively adding weak learners (decision trees) to the ensemble. The objective function consists of two components: a loss function that measures the model's prediction error and a regularisation term that controls the complexity of the model. The formula for the XGBoost objective function can be expressed as Objective Function = Loss Function + Regularization Term. The general form of the objective function for XGBoost is:

$$\sum [L(y_i + \overline{y_i}) + \Omega(f_k)] + \gamma * K$$

where:

$L(y_i + \overline{y_i})$ is the loss function that measures the prediction error between the true label $y_i$and the predicted value $\overline{y_i}$.

$\Omega(f_k)$ represents the regularization term that penalizes complex models by adding up the scores (or weights) of the individual trees in the ensemble. It helps to control overfitting and improve generalization.

γ is a regularization parameter that determines the strength of the regularization term.

K is the number of trees in the ensemble.
Here is how the XGBoost is used in this thesis with the dataset:

- **Data Preparation**: Prepare the dataset with labelled driving behaviour samples. Each sample should have features (e.g., velocity, acceleration) and corresponding labels ("Abnormal" or "Normal").
- **Feature Engineering**: Extract relevant features from the dataset that can help differentiate between Abnormal and Normal driving behaviours.
- **Training and Test Sets**: Separate the data into training and test sets. XGBoost will be trained on the training set, and its performance will be evaluated on the test set.
- **Model Training**: Train an XGBoost model on the training set. Specify the objective function as binary logistic regression, suitable for binary classification tasks. The XGBoost library provides parameters to control various aspects of the model, such as the number of trees, maximum depth of trees, and learning rate.
- **Model Evaluation**: Evaluate the trained XGBoost model on the test set. Calculate evaluation metrics such as accuracy, precision, recall, and F1-score to assess the model's performance in classifying spam and non-spam emails.
- **Fine-tuning**: Adjust the model's hyperparameters, such as the learning rate, maximum depth, or regularization parameters, through cross-validation or grid search to improve the model's performance.
- **Model Deployment**: Deploy the trained XGBoost model in another dataset to classify driving behaviours as Abnormal or Normal in real-time.

Random Forest

Another supervised machine learning model cited in this thesis is Random Forest. The random forest algorithm proposed by (Breiman, 2001) combines several randomized decision trees and aggregates their predictions by averaging. Growing an ensemble of trees and letting them vote for the most popular class has significantly improved classification accuracy. To grow these ensembles, random vectors are often generated that govern each tree's growth in the choir.

For the kth tree, a random vector Θ_k is generated, independent of the past random vectors Θ_1,..., Θ_{k-1} but with the same distribution, and a tree is grown using the training set and Θ_k, resulting in a classifier $h(x, \Theta_k)$ where x is an input vector. The random split selection Θ consists of a number of independent random integers between 1 and K. The nature and dimensionality Θ depend on its use in tree construction. After generating many trees, they vote for the most popular class, and this procedure is called random forests.

A simplified example to illustrate the principles of Random Forest for a binary classification task with the dataset in this thesis:

- **Data Preparation**: In this thesis, we want to detect whether the driving behaviour is normal based on various features such as velocity, acceleration, coordinate, and distance. The dataset contains labelled samples, indicating whether each driving behaviour is abnormal.

- **Ensemble Construction**:
  a. Random Sampling: Randomly select a subset of behaviours from the dataset (with replacement) to create multiple bootstrap samples.
  b. Feature Randomness: Randomly select a subset of features from the available features (velocity, acceleration, coordinate, distance) for each decision tree in the ensemble.
- **Decision Tree Training**: For each bootstrap sample, construct a decision tree using the selected subset of features and the corresponding subset of behaviours. Split the behaviour data based on different feature values to maximize information gain or minimize impurity (e.g., Gini impurity, entropy). The splitting process continues recursively until reaching a stopping criterion, such as a maximum tree depth or the minimum number of samples per leaf.
- **Ensemble Aggregation**: For a new behaviour, each decision tree in the ensemble independently predicts whether the behaviour is abnormal or not based on the selected features. The final prediction is determined by majority voting: the class that receives the most votes across all decision trees is assigned the final predicted class for the behaviour.

5.1.2 Unsupervised ML

With unsupervised learning, machine learning models do not possess example input-output pairs that allow them to learn a function that maps the input features to outputs. Instead, they learn by finding structure within the input features. In unsupervised learning, "structure" refers to patterns, relationships, or regularities within the input features of the data. It involves identifying inherent dependencies or similarities among the data points without using labelled output information. By identifying structure, unsupervised learning models can uncover hidden patterns or groupings in the data, providing a deeper understanding of its inherent properties. Isolation Forest and Robust Covariance are introduced as the two supervised machine learning models.

Isolation Forest

To look for anomalies, Isolation Forest (Lesouple et al., 2021) generates random isolation trees to isolate each data point. The number of branches required to isolate each point is computed for each tree. The mean of this number of branches defines the expected path length, which is used to isolate a point of interest. The expected path length is generally small for anomalies (contrary to nominal data) since anomalies are far from the majority of nominal data.

In statistics, the deviation can be assessed by the Z-score. The generalization of the Z-score for a point x_i in the case of a p-dimensional multi-variate probability distribution with some mean μ and covariance matrix Σ is known as Mahalanobis distance d_i, which is given by:

$$d_i = \sqrt{(x_i - \mu)^T \Sigma^{-1} (x_i - \mu)}$$

Here is a simplified example to illustrate the Isolation Forest algorithm for anomaly detection:

- **Data Preparation**: The dataset comprises various features such as velocity, acceleration, coordinate and distance.
- **Isolation Forest Construction**:
  a. Random Subsampling: Randomly select a subset of instances from the dataset.
  b. Random Feature Split: Randomly select a feature and a split value to partition the selected instances. The split value can be any value within the range of the selected feature.
  c. Recursive Partitioning: Recursively split the instances based on the selected feature and split value. Continue this process until each instance is isolated in a separate leaf node or a predefined stopping criterion is met (e.g., maximum tree depth).
- **Path Length Calculation**: Measure the average path length required to isolate each instance in an isolation tree. The path length is the number of edges traversed from the root to reach a particular instance. Anomalies are expected to have shorter average path lengths compared to normal instances.
- **Anomaly Score Calculation**: Calculate an anomaly score for each instance based on the average path length across all isolation trees. Instances with shorter average path lengths (fewer splits required for isolation) are assigned higher anomaly scores, indicating a higher likelihood of being an anomaly.
- **Anomaly Detection**: Identify instances with anomaly scores above the threshold as anomalies. These instances represent abnormal driving behaviours.

Robust Covariance

The Robust Covariance technique assumes that normal data points have a Gaussian distribution, and accordingly estimates the shape of the joint distribution (i.e., estimates the mean and covariance of the multivariate Gaussian distribution) (Nikita Butakov, 2020). It is based on the fact that outliers lead to an increase of the values (entries) in Σ , making the spread of the data apparently larger. Consequently, $|\Sigma|$(the determinant) will also be larger, which would theoretically decrease by removing extreme events. Rousseeuw and Van Driessen (Peter J.Rousseeuw & Driessen Van Katrien, 1999) developed a computationally efficient algorithm that can yield robust covariance estimates. The method is based on the assumption that at least h out of the n samples are "normal" (h is a hyperparameter). The algorithm starts with k random samples with $(p+1)$ points. For each k sample, μ, Σ, and $|\Sigma|$ are estimated, the distances are calculated and sorted in increasing order, and the h smallest distances are used to update the estimates. In their original publication, the subroutine of computing distances and updating the estimates of μ, Σ, and $|\Sigma|$ is called a "C-step" and two such steps are sufficient to find good candidates (for μ and Σ) among the k random samples. In the next step, a subset of size m with the lowest $|\Sigma|$ (the best candidates) is considered for computation until convergence, and the one estimate whose $|\Sigma|$ is minimal is returned as output.

5.1.3 Semi-supervised

Semi-supervised machine learning is an approach that falls between supervised and unsupervised learning. In semi-supervised learning, the training data consists of labelled examples (data points with assigned labels) and unlabeled examples (data points without assigned labels). Labelled data collection can be challenging and time-consuming, requiring expert annotation or manual labelling. Semi-supervised learning allows us to make the most of the available labelled data by combining it with more unlabeled data, which is typically easier and cheaper. It helps overcome the limited labelled data limitation and enhances the learning process.

Hierarchical Extreme Learning Machines (HELM)

Here, a newly proposed algorithm, the H-ELM algorithm proposed by Tang et al. (2016), is introduced. It is an extension of the ELM algorithm that can be performed with high-speed training, good generalization and universal approximation/classification capability.

The HELM is a feed-forward neural network with multiple hidden layers. It consists of two main steps: unsupervised feature representation and supervised feature classification (L. Chen et al., 2018). In the first step, the HELM aims to learn a sparse encoder in an unsupervised manner, which converts the raw input into higher-level representation. The encoder possesses multiple hidden layers and is trained layer by layer. Given a training set with N samples, say $(X_i, Y_i)(X_i \in R^n, Y_i \in R^t, i = 1,2,3, \dots, N)$, where X_i and Y_i denote the feature representation and the target of the ith sample, respectively. Suppose the encoder consists of K hidden layers, each with $L_i (1 \le i \le K)$ neurons. The output $O = [o_1, o_2, \dots, o_N]^T$ can be expressed as:

$$\sum_{i=1}^{K} \beta_i g(W_i \cdot x_j + b_j) = o_j, j = 1,2, \dots, N$$

where g(x) is the activation function, β_i is the output weight, W_i is the input weight and b_j is the jth bias of the first hidden layer. Ideally, there should be:

$$\sum_{j=1}^{N} \|o_j - Y_j\| = 0$$

that is, there exists β_i, W_i and b_i such that

$$\sum \beta_i g(W_i \cdot x_j + b_j) = Y_j, j = 1,2, \dots, N$$

which can be represented by matrixes as

$$H\beta = Y$$

where H is the output of the hidden layer node, β is the output weight, and Y is the desired output.

$$H(W_1, W_2, \dots, W_K, b_1, b_2, \dots, b_K, x_1, x_2, \dots, x_N) = \begin{bmatrix} g_1(X_1) & \cdots & g_{K_1}(X_1) \\ \vdots & \ddots & \vdots \\ g_1(X_N) & \cdots & g_{K_1}(X_N) \end{bmatrix}$$

To train the single hidden layer ELM neural network is equivalent to obtaining $\hat{\beta}$ such that

$$\|H\hat{\beta} - T\| = \min_{\beta} \|H\beta - T\|$$

When choosing the mean square error (MSE) as the measure, this formula is equivalent to minimizing the following loss function:

$$Loss = \sum_{j=1}^{N} (\sum_{i=1}^{K} \beta_i g(W_i \cdot x_j + b_i) - Y_j)^2$$

The ELM allows the weights β and the deviation between the hidden layer and the inputs to have random values that can be sampled from any distribution. This means that the learning step only determines the optimal weight β between the hidden layer and the output.

The drawback of the pure ELM is that its shallow architecture cannot effectively handle data contents, even with many hidden nodes. HELM, which hierarchically stacks multi-layers of ELM, is one of the most successful attempts to create a deeper structure based on the ELM principles. Therefore, HELM is introduced. In this study, hierarchical ELM layers were first trained using only normal data without any anomalies. The ELMs can capture the most critical input data features by minimising the reconstruction loss. Then the captured features are transferred to the one-class classifier, which is further trained to obtain a threshold using a validation dataset unseen during training. The validation dataset also only contains normal data samples. The model framework related to this thesis is shown in Figure 20.

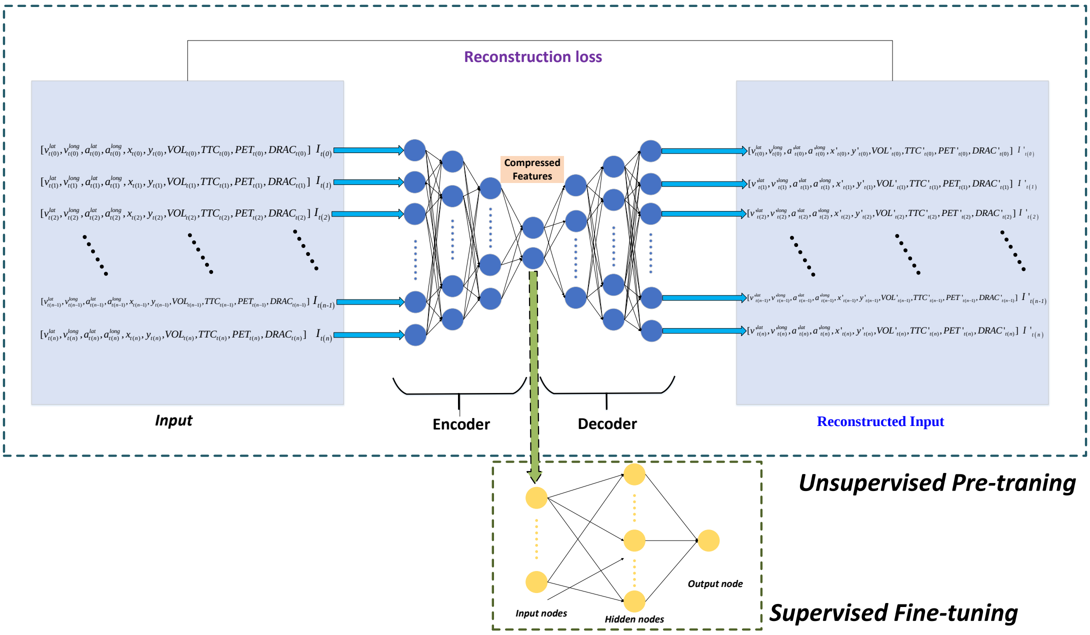

Figure 20 HELM Framework (Drawn by *Zhang Lanxin*)

Input Layer: The HELM's input layer consists of nodes representing each behaviour's features. Each node corresponds to a specific feature.

First Hidden Layer: The first hidden layer captures low-level features. It consists of randomly initialised hidden nodes. Each hidden node connects to the input layer and performs computations using the ELM algorithm.

Second Hidden Layer: The outputs from the first hidden layer serve as inputs to the second hidden layer. This layer captures more complex features and representations built on low-level features. Again, randomly initialised hidden nodes are used.

Output Layer: The output layer categorises the behaviours into the respective classification. It takes the outputs of the second hidden layer as inputs and performs the final classification using a suitable activation function.

5.2 Ablation Study Regarding Features

In this section, a comparative experiment on features is designed. This experiment mainly compares the results of inputting different feature condition models under the same model.

5.2.1 Features Overview

In the literature review, some studies have applied simple Surrogate measures of safety, such as time-to-collision. However, there are two significant shortcomings of the conventional TTC: (1) the scenario is regarded as safe when the speed of the following vehicle is less than or equal to that of the leading vehicle, even though the relative distance could be minimal (Kuang et al., 2015); and (2) the vehicle pair is assumed in the same lane, and only the longitudinal movements are calculated (Xing et al., 2019). To address these limitations of TTC, a new TTC is proposed in this thesis, called two-dimensional TTC:

$$2DTTC = \begin{cases} \frac{DTC}{|v_i - v_j|}, if\ the\ direction\ of\ DTC\ is\ the\ same\ with\ (v_i - v_j) \\ inf, \quad if\ the\ direction\ of\ DTC\ is\ opposite\ with\ (v_i - v_j) \end{cases}$$

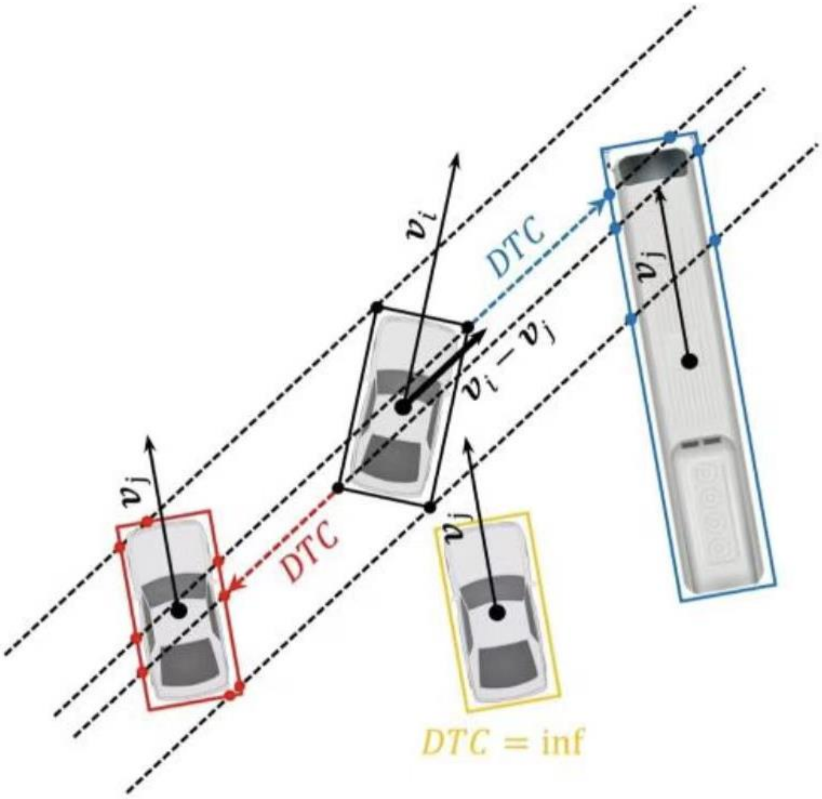

Figure 21 2DTTC scenario (Xing et al., 2019)

In general, only encounters with a minimum TTC of less than 1 s are considered critical and trained observers appear to operate consistently in applying this threshold value (Van Der Horst & Hogema).

5.2.2 Experiment Design

This experiment consists of three feature settings. In set one, the features inherent in the dataset are input: coordinates, velocity, and vehicle angle. Set two adds further calculated acceleration and distance to set one. Sett three inputs time-to-collision additionally.

Table 13 Input features in different settings

Features Setting	Features Input
1	coordinates/velocity/angle
2	coordinates/velocity/angle/acceleration/distance
3	coordinates/velocity/angle/2D time-to-collision

The results of set one are used as a baseline. Further, the results of set two are used to determine whether the model results become better when introducing new features. Set 3 verifies that SMOS has a significant impact on the model.

Here is the process of the experiment:

Experiment Setup

- Dataset: Prepare a dataset containing samples of normal and abnormal driving behaviour.
- Data Preprocessing: Preprocess the dataset to ensure data quality and consistency.
- Split Dataset: Divide the dataset into training and testing sets.

Set One

- Input Features: Use coordinates, velocity, and vehicle angle.
- Model Training: Train the machine learning models (e.g., Random Forest, XGBoost, Robust Covariance, Isolation Forest, HELM) using the training set.
- Model Evaluation: Evaluate the performance of each model using the testing set. Calculate accuracy, precision, recall, F1-score, false positive rate (FPR), and true positive rate (TPR).

Set Two

- Input Features: Add calculated features of acceleration and distance to the features used in Set One.
- Model Training: Retrain the machine learning models using the extended feature and training sets.
- Model Evaluation: Evaluate the updated models using the testing set and calculate the performance metrics.

Set Three

- Input Features: Include the additional feature of 2D time-to-collision and the features used in Set Two.
- Model Training: Retrain the machine learning models with the expanded feature set and the training set.
- Model Evaluation: Evaluate the updated models using the testing set and calculate the performance metrics.

Performance Comparison

- Analyze and compare the performance metrics (accuracy, precision, recall, F1-score, FPR, TPR) for each set and each machine learning model.
- Identify patterns or trends in the models' performance by adding new features in each set.
- Summarize the findings and discuss the effectiveness of the different sets and models in detecting abnormal driving behaviour.

5.3 Model performance measures

Various metrics will be adopted to evaluate the overall performance of the selected model, and the discrimination evaluation of the optimal model can be defined based on the confusion matrix (M & M.N, 2015), as shown in Table 14.

Table 14 Confusion Matrix and the Corresponding Array Representation

	Actual Positive Class	**Actual Negative Class**
Predicted Positive Class	True-positive (TP)	False-negative (FN)
Predicted Negative Class	False positive (FP)	True-negative (TN)

One class is called positive, and the other is called negative in binary classification. Positive classes represent events the model tests for, and negative classes represent other possibilities. For example, the positive class in abnormal driving behaviour detection might be "abnormal." and contrast with the negative class. True-positive (TP) and True-negative (TN) denote the number of positive and negative instances that are correctly classified. In this research, TP represents the number of correctly detected anomalies, and TN represents the number of correctly detected normally. Meanwhile, False-positive (FP) and False-negative (FN) denote the number of misclassified positive and negative instances, which means the number of incorrectly detected anomalies/normal. Then, accuracy, precision and recall were calculated based on the four terms.

Accuracy refers to the proportion of true results among the total number of cases examined.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision is utilized to gauge the accurate prediction of positive patterns among the total predicted patterns in a positive class.

$$Precision = \frac{TP}{TP + FP}$$

Another beneficial measure is recall, which answers a different question: what proportion of actual Positives is correctly classified?

$$Recall = \frac{TP}{TP + FN}$$

The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall.

$$F - score = 2 \times \frac{precision \times recall}{precision + recall}$$

Finally, the Receiver Operating Characteristic-Area Under the Curve (ROC AUC) is introduced to evaluate the model, which determines areas where the evaluated model is classified better within normal and anomaly situations. This curve plots two parameters: True Positive Rate and False Positive Rate.

$$TPR = \frac{TP}{TP + FN}$$
$$FPR = \frac{FP}{FP + TN}$$

5.4 Conclusion

This section introduced this thesis's five machine learning models: XGBoost, Random Forest, Isolation Forest, Robust Covariance, and Hierarchical Extreme Learning Machines (HELM). In addition, a detailed description of the experimental process is explained in this chapter. The feature ablation study compares the results obtained by changing different feature conditions in the model. Finally, various metrics were introduced and will be employed in the next chapter to compare the performance of different models in detecting abnormal driving behaviour. These metrics included accuracy, precision, recall, and F1-score, which will evaluate each model's effectiveness.

6 Results, analysis and discussion

This chapter will address sub-questions 1.2, 1.3, 1.4, 1.5, and 1.6. Regarding sub-question 1.2 and 1.3, the unsupervised and supervised machine learning models as the baseline will be briefly introduced. In contrast, the semi-supervised machine learning model will be elaborated as the innovation point of this research. Then, the results of different types of models will be presented and compared in detail in the next chapter. For sub-question 1.4, 1.5, and 1.6, the same machine learning models will be used, and only change the input conditions to compare the results of different models under different conditions.

6.1 Comparison result for ML models

This section presents three figures representing the results of supervised learning, unsupervised learning, and semi-supervised learning models for abnormal driving behaviour detection. These figures are in the form of matrix plots, commonly used to visualize the outcomes of machine learning algorithms. By examining these figures, the performance of different models in detecting abnormal driving behaviours can be illustrated.

Figure 22 illustrates the results of the supervised learning model. This model is trained using labelled data to learn the patterns and features of abnormal driving behaviours. The matrix plot displays the predicted outcomes for different driving behaviours, providing insights into the model's accuracy and false positive rates in detecting abnormal behaviours.

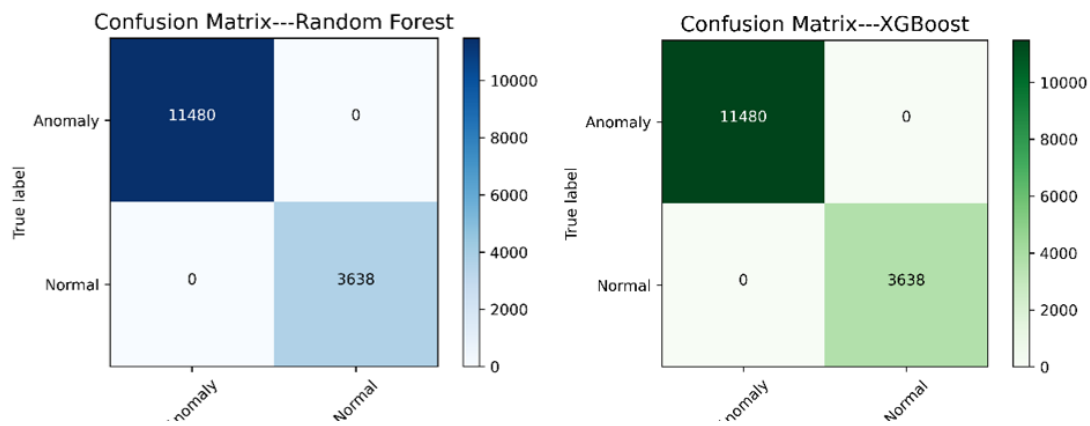

Figure 22 Supervised machine learning result

Figure 23 showcases the results of the unsupervised learning model. This model leverages unlabeled data to discover internal structures and patterns within the data, enabling the detection of abnormal driving behaviours. The matrix plot demonstrates the model's results for driving behaviours and observes whether the model can effectively differentiate abnormal behaviours from normal ones.

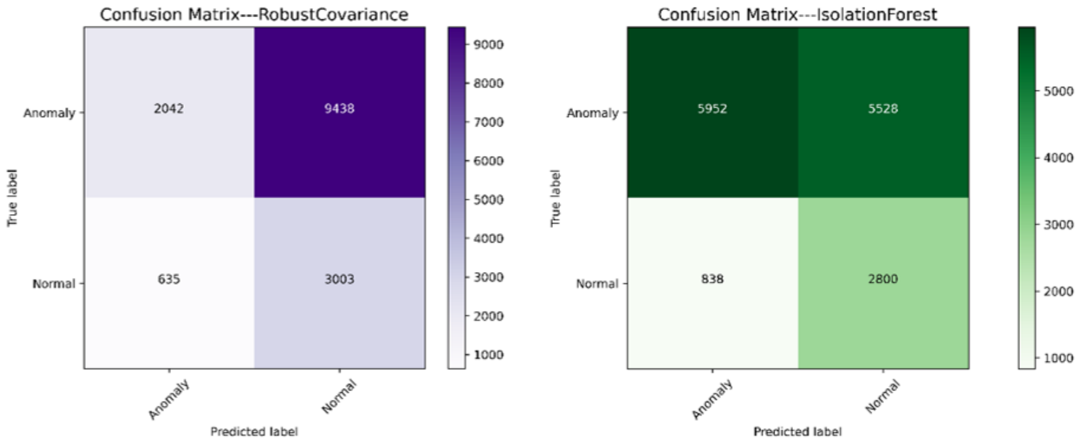

Figure 23 Unsupervised machine learning result

Figure 24 presents the results of the semi-supervised learning model. This model utilizes both limited labelled data and abundant unlabeled data to enhance the detection performance of abnormal driving behaviours. The matrix plot illustrates the model's predictions for abnormal driving behaviours, enabling a comparison of its performance with the supervised and unsupervised models.

The semi-supervised learning model can also distinguish the severity levels of different abnormal driving behaviours, also seen in Figure 24. This means that it not only detects whether a driving behaviour is abnormal but also provides insights into the degree of severity associated with each detected abnormal behaviour.

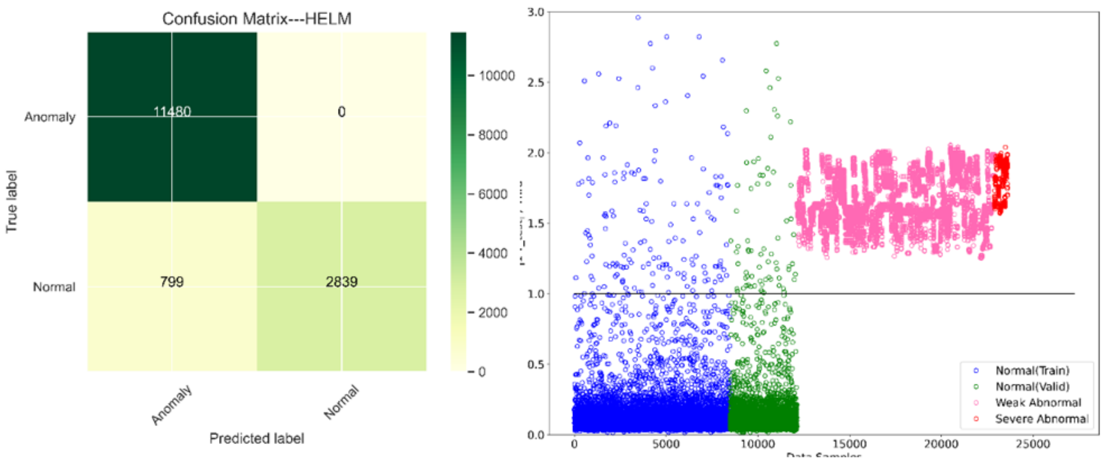

Figure 24 Semi-supervised machine learning result

Table 15 Comparison with different models

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**FPR**	**TPR**

XGBoost	1.000	1.000	1.000	1.000	0.000	1.000
Random Forest	1.000	1.000	1.000	1.000	0.000	1.000
Robust Covariance	0.3337	0.7628	0.1779	0.3735	0.1745	0.1779
Isolation Forest	0.5789	0.8766	0.5185	0.4680	0.2303	0.5185
HELM	0.9471	0.9349	1.0000	0.8766	0.2196	1.0000

Table 15 is a summary of the results of the five models mentioned above.

The XGBoost model demonstrates perfect performance across all metrics, achieving 100% accuracy, precision, recall, and F1-score. It correctly identifies all positive instances without false positives (FPR = 0) and achieves a true positive rate (TPR) 1.000.

Like XGBoost, the Random Forest model achieves excellent performance, with perfect accuracy, precision, recall, and F1-score scores. It also has a false positive rate (FPR) of 0 and a true positive rate (TPR) 1.000.

In supervised learning, the models learn from labelled data, which means they are trained using input-output pairs where the correct output (label) is provided for each input. However, it is important to note that the labelled data might not capture all possible variations of abnormal behaviour. There could be abnormal behaviour that does not follow the predefined rules or patterns captured by the labelled data. In such cases, supervised learning may not achieve 100% accuracy in detecting all types of abnormal behaviour.

The Robust Covariance model performs relatively poorly compared to XGBoost and Random Forest. It has a low accuracy of 0.3337 and lower precision, recall, and F1-score values. The false positive rate (FPR) is 0.1745, indicating a relatively high rate of false positives, while the true positive rate (TPR) is 0.1779.

The Isolation Forest model exhibits moderate performance. It achieves an accuracy of 0.5789 and relatively higher precision and recall values than Robust Covariance. However, the F1-score is lower at 0.4680. The false positive rate (FPR) is 0.2303, indicating a higher rate of false positives, while the true positive rate (TPR) is 0.5185.

The HELM model demonstrates high performance but is slightly lower than XGBoost and Random Forest. It achieves an accuracy of 0.9471, with high precision and recall values. The F1-score is 0.8766, indicating a good balance between precision and recall. The false positive rate (FPR) is 0.2196, while the true positive rate (TPR) is 1.0000.

Below is a detailed analysis of the results of HELM:
Accuracy: The accuracy of the HELM model is 0.9471, which means it correctly classifies approximately 94.71% of the instances in the dataset. It measures overall

correctness and indicates the model's ability to make accurate predictions.

Precision: The precision of the HELM model is 0.9349, indicating that out of all the instances predicted as abnormal driving behaviour, 93.49% are true positives. Precision measures the proportion of correctly predicted positive instances out of all predicted positives, providing insight into the model's ability to avoid false positives.

Recall (True Positive Rate): The recall of the HELM model is 1.0000, meaning it identifies all instances of abnormal driving behaviour in the dataset. It measures the proportion of true positive instances correctly identified by the model out of all actual positives. In other words, the model has a high ability to detect abnormal driving behaviour.

F1-Score: The F1-score of the HELM model is 0.8766, which is the harmonic mean of precision and recall. It provides a balanced measure of the model's performance, considering precision and recall. A higher F1 score indicates a better balance between the two metrics.

False Positive Rate (FPR): The false positive rate of the HELM model is 0.2196, which represents the proportion of instances that are incorrectly classified as abnormal driving behaviour out of all actual negative instances. A lower FPR indicates a lower rate of false alarms or false positives.

At the same time, it must be addressed that due to the operational characteristics of the supervised model (which requires early labelling of all input data), it can save time in real-life or ultra-large datasets. In addition, in the dataset, varying degrees of abnormal driving behaviour were calibrated (see data section), and only the semi-supervised model had the potential to identify it. Based on the above viewpoint, the semi-supervised model is the most suitable for detecting abnormal driving behaviour in real life.

6.2 Result for Ablation Study Regarding Features

As mentioned in the previous section, the supervised model must label all input data independent of the input feature conditions. Therefore, this experiment only compares the results of unsupervised and semi-supervised models.

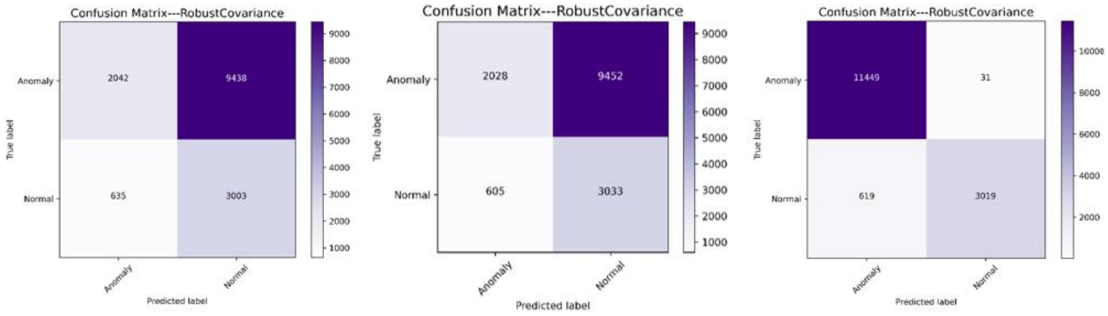

Figure 25 Robust Covariance performance under Set 1,2, and 3

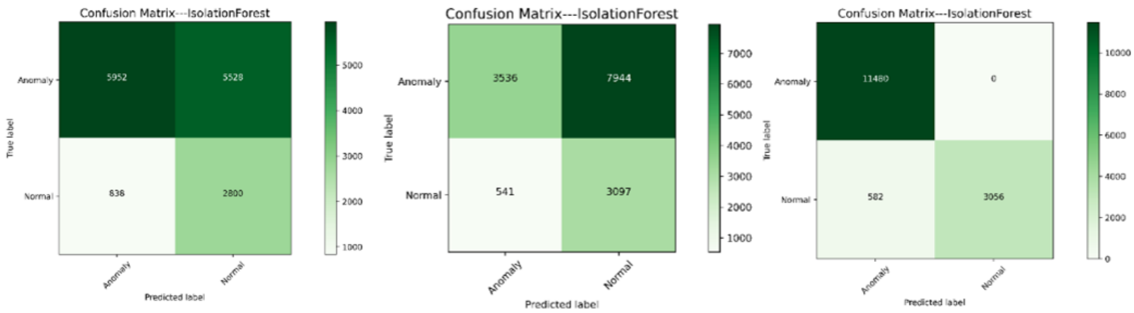

Figure 26 Isolation Forest performance under Set 1,2, and 3

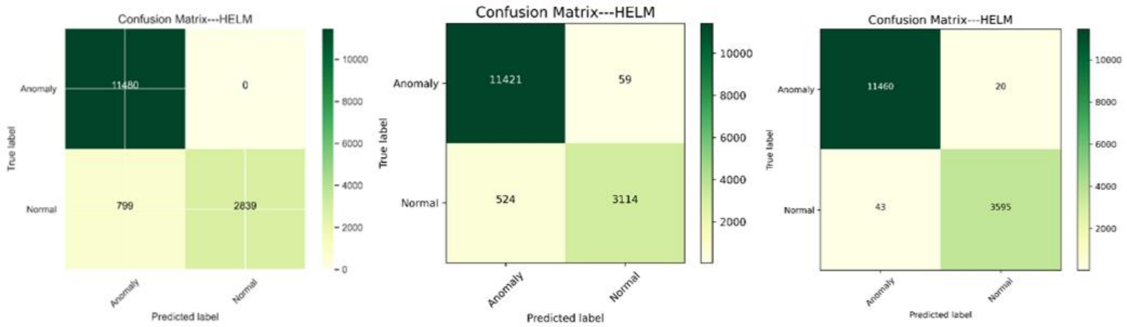

Figure 27 HELM performance under Set 1,2, and 3

Table 16 Comparison results under different set

	SET	Accuracy	Precision	Recall	F1-Score	FPR	TPR
Robust Covariance	1	0.3337	0.7628	0.1779	0.3735	0.1745	0.1779
	2	0.3348	0.7702	0.1767	0.3762	0.1663	0.1767
	3	0.9570	0.9487	0.9973	0.9028	0.1701	0.9973
Isolation Forest	1	0.5789	0.8766	0.5185	0.4680	0.2303	0.5185
	2	0.4387	0.8673	0.3080	0.4219	0.1487	0.3080
	3	0.9615	0.9517	1.0000	0.9131	0.1600	1.0000
HELM	1	0.9471	0.9349	1.0000	0.8766	0.2196	1.0000
	2	0.9614	0.9561	0.9949	0.9144	0.1440	0.9949
	3	0.9958	0.9963	0.9983	0.9913	0.0118	0.9983

Table 17 Input features in different sets

Feature settings	Features Input
1	coordinates/velocity/angle
2	coordinates/velocity/angle/acceleration/distance
3	coordinates/velocity/angle/2D time-to-collision

Set 1: Features Input - coordinates/velocity/angle

- Robust Covariance: The model achieved an accuracy of 1.0, but the precision, recall, and F1-score are relatively low. This indicates that while the model correctly classified some instances, it struggled to identify abnormal driving behaviour accurately.
- Isolation Forest: The model achieved an accuracy of 1.0 and showed improved precision, recall, and F1-score compared to Robust Covariance. It performed better in identifying abnormal driving behaviour based on the given features.
- HELM: The model achieved an accuracy of 0.9471, with high precision, recall, and F1 score. It demonstrated strong performance in detecting abnormal driving behaviour in Set 1.

Set 2: Features Input - coordinates/velocity/angle/acceleration/distance

- Robust Covariance: The model's accuracy slightly increased compared to Set 1, but the precision, recall, and F1-score remained low.
- Isolation Forest: The model's accuracy decreased compared to Set 1, and there was a decline in precision, recall, and F1-score. It struggled to detect abnormal driving behaviour with additional features effectively.
- HELM: The model's accuracy improved to 0.9614, with high precision, recall, and F1 score. It demonstrated better performance in identifying abnormal driving behaviour compared to the other models in Set 2.

Set 3: Features Input - coordinates/velocity/angle/2D time-to-collision

- Robust Covariance: The model's accuracy remained the same as in Set 1, but the precision, recall, and F1-score improved significantly. It better detected abnormal driving behaviour when 2D time-to-collision was considered.
- Isolation Forest: The model's accuracy increased compared to Set 2, and the precision, recall, and F1-score also improved. It demonstrated improved performance with the inclusion of 2D time-to-collision.
- HELM: The model's accuracy improved to 0.9958, with near-perfect precision, recall, and F1-score. It exhibited exceptional performance in identifying abnormal driving behaviour when considering 2D time-to-collision.

Based on the analysis, all three models performed better as more features were added to each feature set. However, HELM consistently outperformed the other models across all scenarios, indicating its effectiveness in detecting abnormal driving behaviour.

Additionally, including 2D time-to-collision as a feature improved the performance of all models, particularly in Setting 3.

6.3 Conclusion

Therefore, in this section, sub-question 1.2, 1.3, 1.4, 1.5, and 1.6 can be answered: **What are the results of machine learning methods for detecting abnormal driving behavior?**

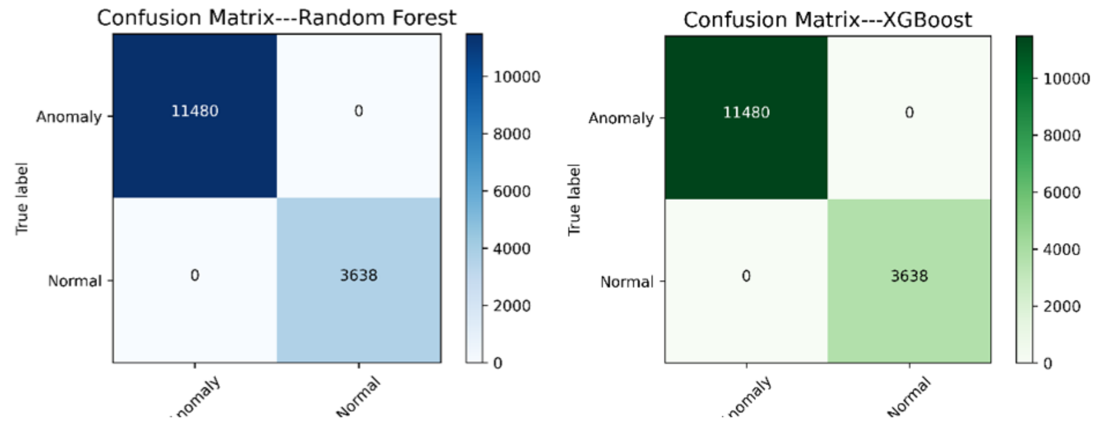

Figure 28 Supervised machine learning result

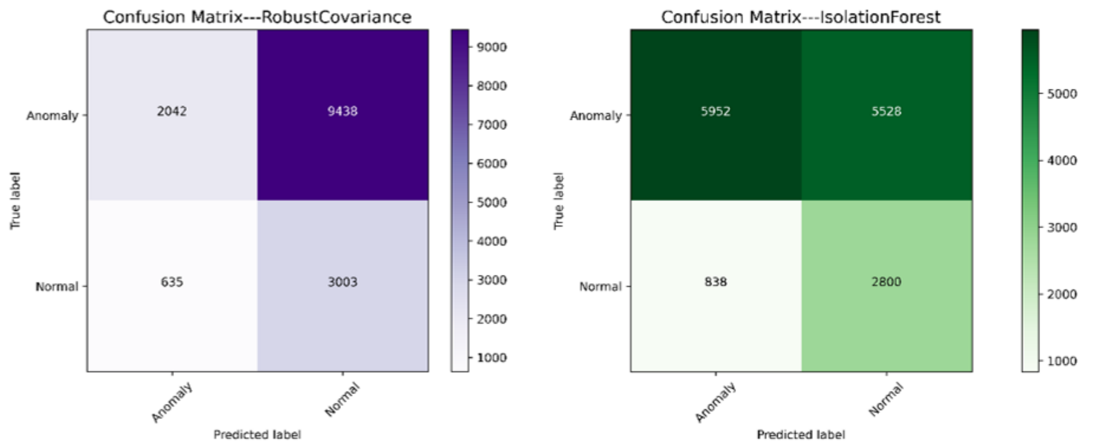

Figure 29 Unsupervised machine learning result

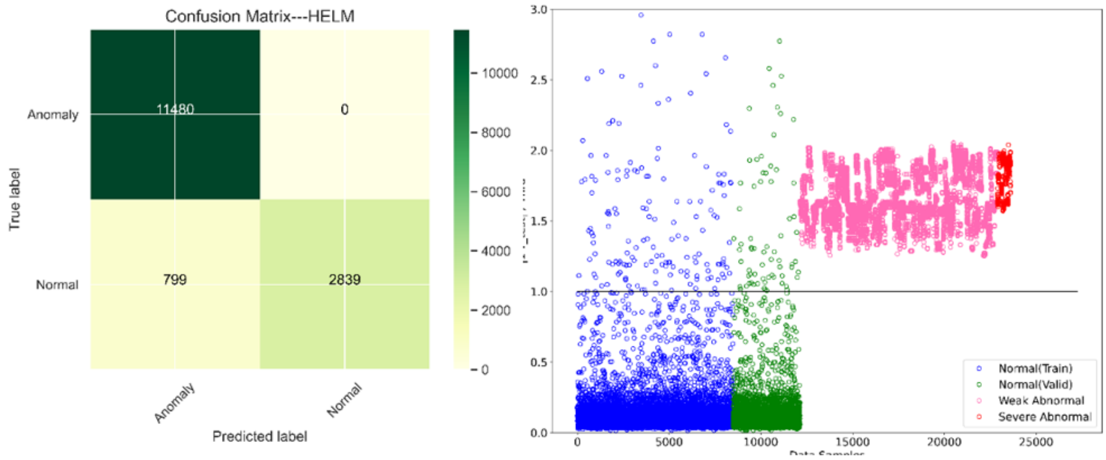

Figure 30 Semi-supervised machine learning result

Table 18 Comparison with different models

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**FPR**	**TPR**
XGBoost	1.000	1.000	1.000	1.000	0.000	1.000
Random Forest	1.000	1.000	1.000	1.000	0.000	1.000
Robust Covariance	0.3337	0.7628	0.1779	0.3735	0.1745	0.1779
Isolation Forest	0.5789	0.8766	0.5185	0.4680	0.2303	0.5185
HELM	0.9471	0.9349	1.0000	0.8766	0.2196	1.0000

What are the differences in the performance of unsupervised, supervised, and semi-supervised machine learning?

This chapter first compares the results of different types of machine learning. It was found that under the same conditions (i.e., without changing any input features), the supervised model performs best. However, the need for a large amount of labelled data in supervised models will increase workloads in real life or when facing massive datasets.

Will SMOS help to improve the performance?

They are incorporating the 2D time-to-collision as a feature enhanced performance for all models, particularly in Setting 3. These findings emphasize the significance of Surrogate Measures of Safety (SMOS) and highlight the potential of HELM in effectively identifying abnormal driving behaviour.

What are the changes in the results of different types of machine learning models after the introduction of SMOS?

Due to the operational characteristics of the supervised model, changing the input features has no practical significance. Therefore, this issue is only discussed for unsupervised and semi-supervised models. After the introduction of SMOS, the performance results of both unsupervised and semi-supervised models have been improved. The unsupervised model has the greatest change in results, with accuracy ranging from less than 50% to over 90%.

Can the model still detect abnormal driving behaviour without using SMOS?

Based on the result above, the Isolation Forest and Robust Covariance models did not detect abnormal driving behaviour without Safety Measures of Safety (SMOS). On the other hand, the semi-supervised HELM (Hierarchical Extreme Learning Machines) model showed promising results even without including SMOS.

7 Conclusions and recommendations

7.1 Conclusions

Differences in the performance of unsupervised, supervised, and semi-supervised machine learning were investigated in this study, shedding light on their respective strengths and limitations. When evaluating the models under the same conditions, the supervised model consistently outperformed the others in accuracy. However, the practical challenges associated with acquiring and labelling a large amount of training data hinder the widespread adoption of supervised learning approaches, especially in real-life scenarios or when dealing with massive datasets.

To address these challenges, semi-supervised learning emerged as a promising alternative. By leveraging a combination of labelled and unlabeled data, semi-supervised models can achieve competitive performance while reducing the dependency on fully labelled datasets， which makes them well-suited for real-world applications and scenarios where labelled data may be limited.

In this research study, integrating Surrogate Measures of Safety (SMOS), particularly incorporating the 2D time-to-collision as a feature, emerged as a pivotal factor in augmenting the performance of the machine learning models. Notably, the inclusion of SMOS yielded substantial improvements in the performance of all models, encompassing both unsupervised and semi-supervised approaches. These compelling findings underscore the significance of SMOS in enhancing the efficacy of abnormal driving behaviour detection across diverse machine-learning models.
The questions raised in this thesis have been resolved, and the following are the answers to the research questions.

The comparative analysis reveals that the supervised model better detects abnormal driving behaviour. However, the practical constraints arising from the requirement of labelled data render the semi-supervised approach more viable in real-world scenarios and when confronted with large-scale datasets. Notably, integrating Surrogate Measures of Safety (SMOS) catalyzes further augmenting the performance of all models, thereby underscoring its pivotal role in effectively identifying and characterizing abnormal driving behaviour. These findings highlight the significance of semi-supervised learning and the utility of SMOS in advancing the field of abnormal driving behaviour detection.

7.2 Reflection

This thesis aims to fill a significant gap in the existing literature on the feature and model aspects of the problem. The previous studies have predominantly overlooked the utilization of Surrogate Measures of Safety (SMOS) as input features, instead focusing on basic motion data such as velocity, angle, and acceleration. However, this thesis takes a novel approach by incorporating SMOS into the analysis and demonstrates its substantial impact on enhancing the model's accuracy. By introducing SMOS as input features, the thesis showcases the potential for improved performance in detecting abnormal driving behaviour.

Moreover, the thesis highlights a disparity in adopting detection methods for abnormal driving behaviour. Previous studies have primarily relied on clustering analysis and shallow machine learning techniques, with limited exploration of semi-supervised models. In contrast, this thesis introduces the Hierarchical Extreme Learning Machines (HELM) as the primary model for abnormal driving behaviour classification. Through empirical analysis, the thesis concludes that HELM offers promising capabilities in accurately classifying abnormal driving behaviour.

However, it is important to note that the dataset used in this thesis encompasses only three types of abnormal driving behaviours. For example, at unsignalized intersections, abnormal driving behaviour is not detected with opposite conflicting. Consequently, the thesis cannot present the complete range of SMOS values (such as Post encroachment time etc.), limiting the comprehensive display of all the SMOS features.

This thesis significantly contributes to the existing literature by emphasizing the importance of incorporating SMOS as input features and leveraging HELM for accurate abnormal driving behaviour detection by addressing these gaps. Nonetheless, future studies should consider expanding the dataset to encompass a broader spectrum of abnormal driving behaviours and associated SMOS to enhance the understanding and detection of such behaviours.

7.3 Recommendation

Firstly, as highlighted in the reflection section of the thesis, it is important to acknowledge that the dataset used in this study may need to be considered better while being the best available at the current stage. Although it provides valuable insights into abnormal driving behaviour, it is limited regarding the types of abnormal driving behaviour included. To further improve the content and analysis of (SMOS), future research must incorporate a more extensive and diverse dataset that encompasses a wide range of abnormal driving behaviours. This will enhance the generalizability of the findings and enable a more comprehensive exploration of SMOS concerning different types of driving behaviour.

In addition, this thesis primarily focuses on investigating the Time to Collision (TTC) as one example of SMOS, neglecting the calculation and analysis of other important SMOS features such as Post Encroachment Time (PET) and various others. While the findings related to TTC are valuable, it is essential to recognise that SMOS encompasses a wide range of measures that provide valuable insights into driving behaviour and safety. Therefore, future directions should emphasise including multiple SMOS features in the analysis to capture a more comprehensive understanding of abnormal driving behaviour.

By incorporating a more diverse dataset and including additional SMOS features beyond TTC, future research in this domain will significantly contribute to the field of abnormal driving behaviour detection. This will enable researchers to delve deeper into analysing various SMOS metrics, improving our understanding of their significance and impact on identifying and classifying abnormal driving behaviour. Additionally, it will provide valuable insights into developing more robust and accurate models for detecting abnormal driving behaviour and improving overall road safety.

Besides refining the dataset, future research should also emphasise the need for more precise and accurate data annotation processes. Improving the quality of data labelling will enhance reliability.

8 Reference

Chen, Z., Yu, J., Zhu, Y., Chen, Y., & Li, M. (2015). D3: Abnormal driving behaviors detection and identification using smartphone sensors. *2015 12th Annual IEEE International Conference on Sensing, Communication, and Networking, SECON 2015*, 524–532. https://doi.org/10.1109/SAHCN.2015.7338354

Dhar, P., Shinde, S., & Bhaduri, A. (2014). Unsafe Driving Detection System using Smartphone as Sensor Platform. In *International Journal of Enhanced Research in Management & Computer Applications* (Vol. 3). www.erpublications.com

Jia, S., Hui, F., Li, S., Zhao, X., & Khattak, A. J. (2020). Long short-term memory and convolutional neural network for abnormal driving behaviour recognition. *IET Intelligent Transport Systems*, *14*(5), 306–312. https://doi.org/10.1049/iet-its.2019.0200

Kim, D. G., Lee, C., & Park, B. J. (2016). Use of digital tachograph data to provide traffic safety education and evaluate effects on bus driver behavior. *Transportation Research Record*, *2585*, 77–84. https://doi.org/10.3141/2585-09

Lim, S., & Yang, J. H. (2016). Driver state estimation by convolutional neural network using multimodal sensor data. *Electronics Letters*, *52*(17), 1495–1497. https://doi.org/10.1049/el.2016.1393

M, H., & M.N, S. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 01–11. https://doi.org/10.5121/ijdkp.2015.5201

Nikita Butakov. (2020). *How to build robust anomaly detectors with machine learning*.

Peter J.Rousseeuw, & Driessen Van Katrien. (1999). *A_fast_algorithm_for_the_minim*.

Planek, T. W., Sinelnikov, S., Thomas, J., Kolosh, K., & Porretta, K. (2015). Letter from the Editors - Fourth international symposium on naturalistic driving research. In *Journal of Safety Research* (Vol. 54, p. 29). Elsevier Ltd. https://doi.org/10.1016/j.jsr.2015.06.003

Xing, L., He, J., Abdel-Aty, M., Cai, Q., Li, Y., & Zheng, O. (2019). Examining traffic conflicts of up stream toll plaza area using vehicles' trajectory data. *Accident Analysis and Prevention*, *125*, 174–187. https://doi.org/10.1016/j.aap.2019.01.034

Zheng, O., Abdel-Aty, M., Yue, L., Abdelraouf, A., Wang, Z., & Mahmoud, N. (2022). *CitySim: A Drone-Based Vehicle Trajectory Dataset for Safety Oriented Research and Digital Twins Figure 1. Post encroachment time conflicts in a single frame from the CitySim dataset Expressway A weaving segment location*. https://github.com/ozheng1993/UCF-SST-CitySim-Dataset.