Context-Aware Intelligent Decisions: Online Assessment of Heavy Goods Vehicle Driving Risk

Jimiama Mosima Mafeni Mase

Thesis submitted to The University of Nottingham
for the Degree of Doctor of Philosophy
in Computer Science

May 16, 2023

Publications

This section presents a list of publications formed as part of the research work for this thesis.

1. Jimiama M. Mase, Peter Chapman, Christian Wagner, and Grazziela P. Figueredo, "Driving Risk Assessment Using Intervals and Weighted Fuzzy Rules". To be submitted.

2. Jimiama M. Mase, Peter Chapman, and Grazziela P. Figueredo, "A Review of Intelligent Systems for Driving Risk Assessment". Submitted.

3. Kumar Aayush, Jimiama M. Mase, Divish Rengasamy, Benjamin Rothwell, Mercedes T. Torres, David A. Winkler, and Grazziela P. Figueredo, "EFI: A Toolbox for Feature Importance Fusion and Interpretation in Python". Published.

4. Jimiama M. Mase, Natalie Leesakul, Grazziela P. Figueredo, and Mercedes T. Torres, "Facial identity protection using deep learning technologies: an application in affective computing". Published.

5. Divish Rengasamy, Jimiama M. Mase, Aayush Kumar, Benjamin Rothwell, Mercedes T. Torres, Morgan R. Alexander, David A. Winkler, and Grazziela P. Figueredo, "Feature importance in machine learning models: A fuzzy information fusion approach". Published.

6. Jimiama M. Mase, Peter Chapman, Grazziela P. Figueredo, and Mercedes T. Torres, "Benchmarking deep learning models for driver distraction detection". Published.

7. Jimiama M. Mase, U. Agrawal, D. Pekaslan, M. Mesgarpour, P. Chapman, M. T. Torres, and Grazziela P. Figueredo, "Capturing uncertainty in heavy goods vehicle driving behaviour". Published.

8. Jimiama M. Mase, Shazmin Majid, Mohammad Mesgarpour, Mercedes T. Torres, Grazziela P. Figueredo, and Peter Chapman, "Evaluating the impact of heavy goods vehicle driver monitoring and coaching to reduce risky behaviour". Published.

9. Jimiama M. Mase, Peter Chapman, Grazziela P. Figueredo, and Mercedes T. Torres, "A hybrid deep learning approach for driver distraction detection". Published.

10. Utkarsh Agrawal, Jimiama M. Mase, Grazziela P. Figueredo, Christian Wagner, Mohammad Mesgarpour, and Robert I. John, "Towards real-time heavy goods vehicle driving behaviour classification in the united kingdom". Published.

Abstract

There is a growing interest in assessing the impact of drivers' actions and behaviours on road safety due to the numerous road fatalities and costs attributed to them. For Heavy Goods Vehicle (HGV) drivers, assessing the road safety risks of their behaviours is a subject of interest for researchers, governments and transport companies, as nations rely on HGVs for the delivery of goods and services. However, HGV driving is a complex, dynamic, uncertain and multifaceted task, mostly influenced by individual traits and external contextual factors. Advanced computational and artificial intelligence (AI) methods have provided promising solutions to automatically characterise the manner by which drivers operate vehicle controls and assess their impact on road safety. However, several challenges and limitations are faced by the current intelligence-supported driving risk assessment approaches proposed by researchers, such as: (1) the lack of comprehensive driving risk datasets; (2) information about the impact of inevitable contextual factors on HGV drivers' responses is not considered, such as drivers' physical and mental states, weather conditions, traffic conditions, road geometry, road types, and work schedules; (3) ambiguity in the definition of driving behaviours is not considered; and (4) imprecision of AI models, and variability in experts' subjective views are not considered.

To overcome the aforementioned challenges and limitations, this multidisciplinary research aims at exploring multiple sources of data including information about the impact of contextual factors captured from crucial stakeholders in the HGV sector to develop a reliable context-aware driving risk assessment framework. To achieve this aim, AI methods are explored to accurately detect drivers' driving styles, affective states and driving postures using telematics data, facial images, and driver posture images respectively. Subsequently, due to the lack of comprehensive driving risk datasets, fuzzy expert systems (FESs) are explored to fuse detected driving behaviours and perceived external factors using knowledge from domain experts.

The key findings of this research are: (1) recurrent neural networks are effective in capturing the temporal dynamics and differences between the different types of driver distraction postures and affective states; (2) there is a trade-off between efficiency and privacy in processing facial images using AI approaches; (3) the fusion of driver behaviours and external factors using FESs produces realistic, reliable and fair driving risk assessments; and (4) a hierarchical representation of a decision-making process simplifies reasoning compared to flat representations.

Acknowledgements

I would like to thank my supervisor, Dr Grazziela Figueredo, who has supported, guided and mentored me throughout my PhD with patience and leadership. She contributed to my remarkable track record of publications in multiple conferences and journals by giving me intellectual freedom to research on areas and technologies of my interest. She also introduced me to amazing researchers for collaboration and networking, such as Dr Divish Rengasamy and Dr Utkarsh Agrawal, whom have been a great support in brainstorming ideas. Additionally, I would like to thank my entire supervision team, Professor Christian Wagner, Dr Mercedes Torres Torres and Dr Peter Chapman, who were always available to support and advice me during the four years of my PhD.

Furthermore, I would like to thank the Center of Doctoral Training and Microlise for funding my PhD and providing the necessary resources to complete my PhD. Specifically, I would like to thank Monica Cano Gomez, Andrea Haworth, Laura Brian and Adrian Hazzard for their endless effort and love in supporting their scholars.

A big thank you to all my friends in my cohort. It was not an easy journey, but I enjoyed every moment with you and thank you for all your support and love. I pray that we all end up where our hearts desire.

Finally, I would like to thank my family and friends. To my parents, I say thank you for the encouragement and frequent checkups. To my wife, Chimamaka, who stood by me in my darkest and brightest moments during my PhD. To my brothers, Carl, Francis, Daniel, Tagem, Lwanga, Delavane, Neville, Kwabenna and Randolph, I thank you all. To my sisters, Juliette, Eli, Nnezi, and Nina, I thank you. To all those who supported me in one way or the other, I thank you.

Contents

List of Figures

List of Tables

List of Acronyms

AI Artificial Intelligence

HGV Heavy Goods Vehicle

UK United Kingdom

ML Machine Learning

FS Fuzzy Set

T1 FS Type 1 Fuzzy Set

GT2 FS General Type 2 Fuzzy Set

MF Membership Function

FLS Fuzzy Logic System

FES Fuzzy Expert System

IAA Interval Agreement Approach

SIA Stakeholder-supported Intelligent Driving Assessment

SIFA Stakeholder-supported Intelligent Fuzzy Driving Assessment

CIDA Current Intelligent Driving Assessment

Chapter 1

Introduction

1.1 Motivation

Globally, Heavy Goods Vehicles (HGVs) are at the forefront of trade and commerce, as both private and public sectors rely on them for the delivery of goods and services. Within the United Kingdom (UK), billions of tonnes of freight are being transported by HGVs annually [4]. As a result of the importance of HGVs to a nation's economy, there are great efforts by researchers, governments and transport companies to reduce the injuries, fatalities and costs associated to HGV driving incidents. The main causes of HGV driving incidents are attributed to risky driving behaviours, such as fatigue, road rage (i.e., angry behaviour), distracted driving, recklessness and aggressive driving [5, 6, 7], as well as environmental conditions, such as harsh weather, high traffic congestion and road quality [8, 9, 10, 11, 12].

With the abundance of sensors installed in HGVs e.g., GPS, driver-facing video cameras, road-facing video cameras, steering sensors, braking sensors etc, computational intelligence approaches are being explored to automatically process data from the sensors, and provide accurate and timely support to drivers about their driving behaviours, environment and risks; or/and provide reliable information to key stakeholders for the effective management of the transport network. The current literature on intelligence-supported driving risk assessment are limited to the manner by which drivers operate vehicle controls [13, 14, 15, 16, 17], and do not consider the impact of contextual factors on driving performance and risks, such as drivers' physical and mental states, weather conditions, traffic conditions, road geometry, road types, and work schedules. Therefore, current approaches could potentially produce incomplete, unrealistic and unfair assessments. For example, consider two HGV drivers (e.g. Bob and Alice) having exactly the same number of driving incidents. Bob was driving in favourable conditions e.g., good weather and no time pressure for delivery, while Alice was driving under poor weather conditions with pressure to deliver on time. This thesis argues that the driving performance and risks in both cases are not the same and that context should be taken into account.

Major challenges also exist in the analysis of driving risk, such as: (1) lack of comprehensive driving datasets; (2) difficulty to detect the different facets of driver behaviour from single sources of data; (3) imprecision of AI techniques; (4) ambiguity in the definition of driving behaviours; (5) variability in stakeholders' views about the influence of external conditions; and (6) lack of data about the synergy between driver traits and external conditions. This multidisciplinary research explores state-of-the-art AI methods to accurately detect the different facets of driver behaviour from multiple data sources i.e., telematics incident data, driver posture images and facial images. Subsequently, it examines Fuzzy Logic Systems (FLSs) [18] to handle the uncertainties and ambiguity in information, and fuse heterogeneous information using knowledge from domain experts.

1.2 Aims, Research Questions and Objectives

The initial aim of this research was to develop and evaluate an end-to-end context-aware intelligent driving assessment system that provides real-time assessments of HGV driving risk by processing multi-modal data streams describing driving behaviours and external factors (i.e., online HGV driving risk assessment). However, due to data privacy constraints, multi-modal data streams capturing the different facets of driver behaviour and external conditions could not be obtained. As result, the aim of this research was adapted to develop AI models that can accurately detect HGV driving behaviours, and develop an intelligent system that can automatically assess the impact of detected driving behaviours on road safety taking into consideration their synergy with external conditions.

Figure 1.1 illustrates the aims of this research. First, state-of-the-art AI models are explored to improve the detection of drivers' driving styles, distraction postures and affective states using telematic driving incident data, driving posture images and facial images respectively. Secondly, a heterogeneous information fusion approach is developed to assess the impact of HGV driving on road safety by fusing detected driving behaviours and perceived external factors.

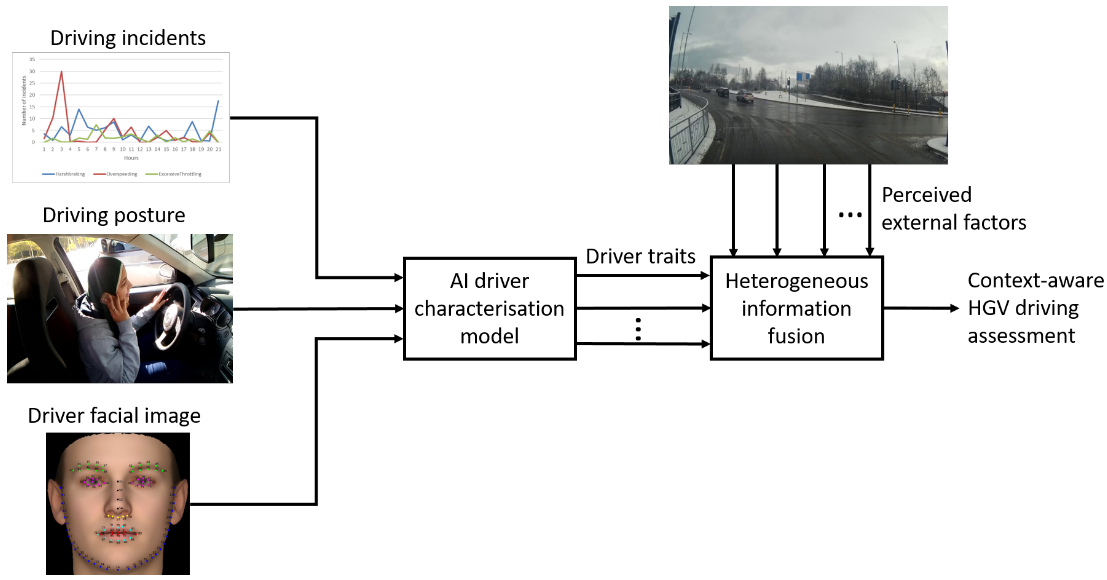


Figure 1.1: A diagram showing the aim of this research.

This thesis plans to answer the following research questions:

- How can the accuracy and reliability of detecting HGV driver behaviour be improved taking into consideration its multifaceted nature? Proper analysis of HGV driving risk cannot be achieved without accurate and reliable information about drivers' actions and behaviours obtained from AI models.

- How can the privacy of drivers be protected when processing data that can easily expose their identities, such as, driver footage data?

- How can a reliable driving risk assessment system that considers the real-world characteristics of the driving environment be developed, taking into consideration the lack of comprehensive driving risk datasets?

- How can the reliability and effectiveness of the driving risk assessment system be evaluated taking into consideration the lack of multi-modal data?

In order to achieve the aims of this research and answer the research questions, the following objectives are developed:

1. Identify the driving behaviours that could potentially affect HGV driving risk by reviewing the main literature on driver behaviour analysis. Subsequently, available data sources are explored using computational and artificial intelligence techniques to develop an intelligent multifaceted driver characterisation framework that can accurately detect the different driving behaviours. The framework should be able to safeguard the identity and privacy of drivers. Achieving this objective will ensure that the information about the detected driving behaviours, fed into the intelligent driving risk assessment system, are accurate and reliable.

2. Identify the most relevant HGV external factors and capture their impact on road safety. As earlier mentioned, the HGV driving environment consist of external factors that influence drivers' responses, but are mostly not captured in available driver data. In this objective, stakeholders in the HGV community (i.e., HGV drivers, HGV transport managers, road safety officers and road safety researchers); who possess a deep understanding of the highly complex and dynamic driving environment; are approached to identify, capture and understand the impact of external contextual factors on HGV driving performance and risk.

3. Capture the synergy between the driving risk factors identified in Objective 2 for the development of a reliable intelligent information fusion system. In this objective, fuzzy logic systems (FLSs) are explored due to their effectiveness to model and embed knowledge from experts using IF/THEN rules [19]. FLSs have shown remarkable performance in several domains, such as, climate classification [20], education [21], healthcare [22, 23], manufacturing [24], and agriculture [25].

4. Evaluate the reliability and effectiveness of the resulting intelligent driving risk assessment system developed in Objective 3. A reliable and effective approach of assessing HGV driving performance and risk will enable trust, acceptance

and adoption. Due to the lack of multi-modal driving risk data, user stories representing realistic HGV driving scenarios are developed with the help of HGV drivers. The assessments of the driving scenarios produced by the system developed in Objective 3 is compared with the assessments produced by HGV drivers and current intelligent driving assessment approaches.

1.3 Thesis Contributions

The thesis proposes three main contributions, as follows:

1. Novel AI techniques to improve the accuracy of detecting driving behaviours and to protect the identity of drivers when processing facial images. A more detailed description of the AI techniques and their components are presented and discussed in Chapter 3.

2. A new framework called stakeholder-supported intelligent driving assessment (SIA) that provides a systematic and collaborative approach to capture and embed contextual information into data-driven decisions. Knowledge from domain experts about the impact of contextual factors are captured, aggregated, visualised and fused to produce fairer and more realistic decisions. The framework is described in Chapter 4.

3. An extension of SIA called stakeholder-supported intelligent fuzzy driving assessment (SIFA) that considers the real-world interactions and uncertainties of risk factors. SIFA enables the capture of the combined effects and synergy between risk factors using fuzzy IF/THEN rules and develops a stakeholder-supported hierarchical rule-based fuzzy inference system. The resulting stakeholder-supported hierarchical fuzzy system improves the interpretability and reliability of assessments compared to SIA and current data-driven driving risk assessment approaches. SIFA is described and evaluated in Chapter 5.

1.4 Thesis Structure

The thesis structure is designed according to the different objectives presented in Section 1.2.

- **Chapter 2** presents relevant background information about driving behaviours and reviews AI techniques employed to characterise driving behaviours and assess their impact on driving risk. The chapter begins by reviewing the main psychological theories on driver behaviour to aid in the understanding and definition of the dynamic and complex HGV driving environment. Subsequently, it reviews the literature on data-driven driver characterisation using AI methods as the reliability of an online HGV driving assessment depends on the accuracy of detected driving behaviours. Most importantly, it reviews the literature on driving risk assessment, which consist of post-hoc analysis of driving risk and AI detection of driving risk. Publication 2 is related to this chapter.

- **Chapter 3** explores state-of-the-art AI techniques for detecting driving behaviours. The chapter introduces and evaluates new AI methods that improve the detection and privacy of driving behaviour using the available datasets. Publications 4, 6, 7, 9 and 10 are related to this chapter.

- **Chapter 4** introduces a systematic and stakeholder-supported framework to identify, capture and embed the impact of contextual factors into the assessment of HGV driving performance and risk. First, the framework engages with a wide variety of stakeholders in the HGV community to capture the impact of contextual factors on road safety. The information obtained from the stakeholders is modelled and aggregated to provide a clear and collaborative representation of the impact of factors. The last stage of the framework proposes an integration solution to embed the contextual information into the assessment of HGV driving. The framework is evaluated using user stories due to the absence of contextual driving risk data. Publications 3, 5 and 8 are related to this chapter.

- **Chapter 5** extends the framework introduced in Chapter 4 to consider the synergistic effects and interactions between contextual factors. The new framework engages with stakeholders in the HGV sector to capture the interactions of contextual factors and utilise stakeholder inputs in the development of a hierarchical rule-based fuzzy inference system. The fuzzy inference system fuses heterogeneous information about driving behaviours and external conditions to produce the likelihood of driving scenarios belonging to specific driving risk categories. The chapter concludes by evaluating the reliability and effectiveness of the framework using user stories of HGV driving scenarios. Publication 1 is related to this chapter.

- **Chapter 6** concludes this thesis by reflecting on the objectives, contributions and findings of this research. It also presents the limitations and future directions of the research.

Chapter 2

Literature Review

2.1 Introduction

This thesis focuses on the development of an online, intelligent and context-aware driving risk assessment framework. First, it investigates artificial intelligence (AI) methods to increase the accuracy and reliability of characterising the main driving behaviours that impact road safety. The research is conducted with particular attention to commercial Heavy Goods Vehicles (HGVs) due to the high costs, fatalities and injuries associated with them. Subsequently, it explores systematic and computational approaches to engage with crucial stakeholders in the HGV sector; capture and embed their expert knowledge about the impact of perceived driver traits and external factors on road safety into the data-driven assessment of HGV driving performance and risk.

This chapter starts by introducing the literature concerning the main psychological theories of driver behaviour in Section 2.2.1. The objective is to aid in the understanding and definition of the dynamic and complex real-world commercial driving environment. The outcome of this section is a list of requirements obtained from the theories for modelling a real-world commercial driving environment, which will be the basis of our investigations and research in this thesis. Later, Section 2.2.2 reviews the literature on driver behaviour analysis based on data to identify and define the main types of behavioural traits that interact to impact road safety. Section 2.2.2 concludes by reviewing data captured by in-vehicle technologies and sensors to provide a list of possible data sources for detecting driver traits in an online driving risk assessment system.

Subsequently, Section 2.3 reviews intelligent methods explored in the literature to automatically characterise driving behaviours. This is important in identifying opportunities to improve HGV driver characterisation as the reliability of the driving risk assessment system depends on the accuracy and reliability of the driving behaviours detected. Strategies explored in the literature to assess driving risk are reviewed in Section 2.4. Finally, Section 2.5 provides a summary of the chapter.

2.2 The Real-World Driving Environment

In this section, the main theories of driver behaviour from a psychological point of view are reviewed. This is to assist in understanding and identifying the main features and characteristics of a real-world driving environment. Subsequently, the literature on driver behaviour analysis based on data is reviewed to identify and explain the main types of driver traits that interact to impact driving risk.

2.2.1 Theories of driver behaviour

In order to model driver behaviour in a real-world driving environment and identify the factors that affect driving risk, three main psychological theories on driver behaviour are reviewed i.e. Theory of Planned Behaviour (TPB) [1, 26, 27], Risk Homeostasis Theory (RHT) [28, 29, 30] and Multiple Resource Theory (MRT) [31, 32, 33, 34]. These theories have been widely and successfully used to understand driving behaviours and environments in traffic safety research.

(a) Theory of planned behaviour

According to the TPB (Figure 2.1), people's attitude towards their behaviour, their subjective norm, and their perceived behavioural control influence their intentions to perform a particular behaviour [1].

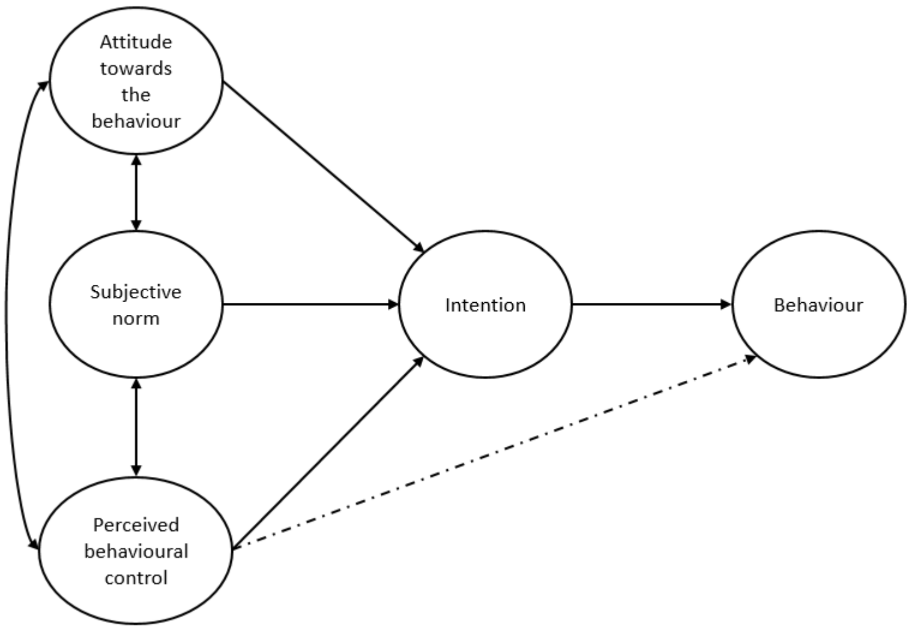


Figure 2.1: The theory of planned behaviour adapted from Ajzan [1]

The attitude towards a behaviour is determined by individual beliefs about the likely consequences of the behaviour; subjective norm is determined by their beliefs about the normative expectations of others; and perceived behavioural control is determined by their beliefs about factors that may facilitate or obstruct the behaviour. Their intentions are defined as "their willingness to try to perform the behaviour" and the behaviour refers to a defined action. For behaviours over which people have incomplete voluntary control, such as control due to external factors, as with commercial vehicles, it is also useful to consider perceived behavioural control in addition to intention. An example of work done using TPB to understand driving behaviours is introduced by Warner and Åberg [35] who investigated driver behaviour before and after the activation of an Intelligent Speed Adaptation (ISA) warning system. Their study demonstrates that TPB can be used as a frame of reference to predict drivers' self-reported speeding behaviour. They claim that drivers decide on a target behaviour (e.g. aggressive or calm driving) and in living up to this decision they continuously monitor their target behaviour.

To summarise, the TPB suggests that the real-world driving environment is a complex and challenging environment, in which drivers' intentions and behaviours are constantly affected by external factors, such as weather, traffic etc. These factors may facilitate or obstruct drivers' intentions and may even change their behaviours. This summary provides the first two characteristics of a real-world driving environment i.e. (1) drivers develop preconceived risk-taking behaviour, which are determined by their objectives and responsibilities; and (2) external factors constantly interact with drivers' behaviours.

(b) Risk homeostasis theory

RHT is a risk compensation model, which states that in any activity, a person accepts a certain level of risk to their safety or safety of other things they value, in exchange for the rewards they hope to get from the activity [36]. In commercial driving, drivers continuously compare and balance the amount of risk they are exposed to, with the amount of risk they are willing to accept [37]. For example, increasing driving speed leads to a higher chance of an incident happening, but also reduces travel time and increases profit. Therefore, there is a trade-off between safety or cost and rewards. If the level of risk they experience is lower than their acceptable risk level, they engage in actions that increase the experienced risk. However, if the level of experienced risk is higher, they exercise greater caution to reduce exposure to risk. The experienced and acceptable risk levels are subjective, as some drivers are prone to take more risks than others (e.g. drivers with aggressive driving styles) based on their attitudes, subjective norms and perceived behavioural control [14, 38]. For example, a driver might set a higher level of risk to deliver their goods on time if their company rewards them for high number of deliveries.

To summarise, RHT describes the constant conflict between driving risk and reward; drivers always accept a certain level of risk depending on the reward. This theory introduces the third characteristic of a real-world driving environment, i.e., there is always some level of risk, which varies with drivers and scenarios depending on the perceived level of reward during the journey. Hence, a driver's level of risk is not constant. This is another reason why it is important to develop online, continuously updated, risk assessments rather than post hoc assessments, to capture changes in drivers' levels of risk.

(c) Multiple resource theory

MRT [39] is a theory of multiple task performance, that tries to answer the research question: 'how is the primary task affected by multiple secondary tasks?'. In driving, MRT describes how secondary tasks e.g. operating a device, looking at road signs interfere with the primary task (driving). MRT was developed to improve human operators' performance in high workload multi-task environments and to account for variability in task interference. It ascertains that secondary tasks requiring the same resources as the driving task produce mutual interference and have greater impact on driving compared to secondary tasks that require different resources as the driving task [40]. For example, an in-vehicle interactive device with touch-control functionalities, requires similar resources (visual and manual) from the driver compared to those required during driving. Wickens [39] introduces four dimensions for describing the resources required by secondary tasks: stages, perception modalities, processing codes, and responses.

- Stages — The resources used for perception, cognition and responding to stimuli. This dimension summarises the entire process of completing a task.
- Perception modalities — This classifies the perception stage into the auditory and visual channels. In this dimension, more complex perception modalities can be included, such as olfactory channels.
- Processing codes — This dimension classifies cognition resources into spatial and verbal processes.
- Responses - The output of the processing dimension leads to the response dimension, which can either be manual or vocal. Manual responses are usually spatial in nature (e.g. steering, joystick or pedals) and the vocal responses are verbal in nature (e.g. speaking).

MRT suggests that the performance of the primary task of driving is determined by the interaction among multiple secondary tasks i.e. driving behaviours. The effect of the co-occurrence and interaction of the behaviours on driving risk depends on their perception modalities and responses. Driving behaviours that require the same resources as the driving task have greater impact on driving risk. This theory introduces another characteristic of driver behaviour i.e. the primary task of driving is influenced by multiple secondary driving behaviours, such as listening to a radio, looking at road signs, and emotions etc.

The characteristics obtained from the psychological theories of driver behaviour are combined to provide a theoretical framework of driver behaviour in the real-world driving environment, presented in Figure 2.2. Using the theories, driver behaviour can be defined as "*a set of driver actions and states that interact concurrently with external and environmental factors during driving to produce some level of driving risk*". It is important to mention that this framework simplifies the complex and dynamic driving environment into more manageable characteristics, in order to create a feasible scope for the thesis. The driver module is constantly being affected by external factors in the driving environment as described by TPB, such as weather conditions, road types and traffic conditions, and by driving rewards, such as compensation to deliver on time and compensation to complete multiple jobs. The driver module consist of a set of driver actions and states that constantly interact with each other and concurrently with the external factors to determine the level of driving risk as described by RHT. Examples of driver actions or responses are accelerating, braking, speeding, and steering, while driver states are the mental, affective and physical states of the driver when driving. The driver states represent the secondary tasks described in MRT. The mental states of drivers are related to their mental workload, decision making, situation awareness and memory e.g., driving confidence, error and lapses [41]. The affective states represent the mood or feelings of the driver such as happy, frustrated, sad and angry. The physical states include behaviours relating to the driver's body e.g., driving postures, eye movements, head movements and physiological states.

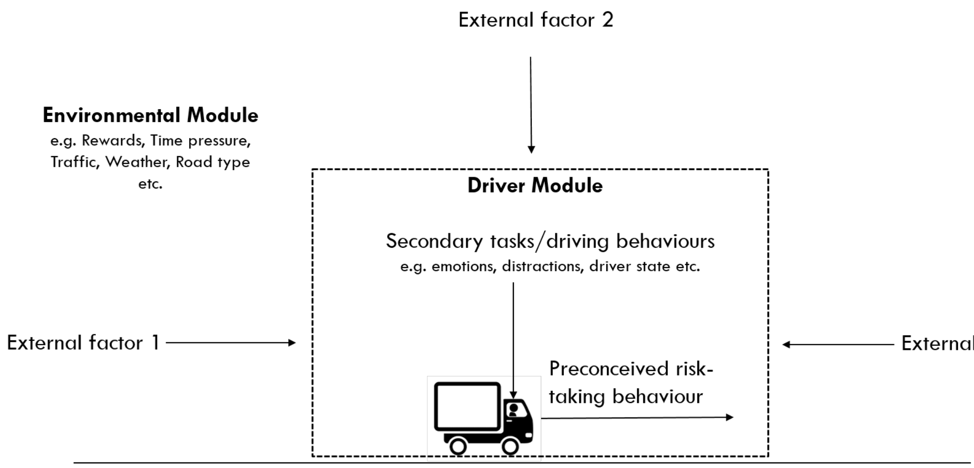


Figure 2.2: The characteristics of the real-world commercial driving environment conceived from the theory of planned behaviour, risk homeostasis theory and multiple resource theory.

2.2.2 Main types of driving behaviours

In this section, the literature on driver behaviour analysis is reviewed to identify the main types of driver behavioural traits that impact driving performance and road safety. Subsequently, the section reviews possible data sources explored by researchers for analysing the different driving behaviours.

After reviewing the main literature on driver behaviour analysis, three main types of driving behaviours are identified: (1) driving postures [42, 43, 44, 45, 46], (2) responses to affective states, e.g., fatigue [47, 48, 49, 49, 50, 51, 52, 53] and emotion [54, 55, 56], and (3) driving styles [57, 58, 59, 60, 61]. These behaviours are described as follows:

- Driving postures — Distracted driving postures occur when a driver fails to attend to the primary task of driving or performs activities while driving that could impact safe driving [62, 63] e.g. phone usage, operating in-vehicle technologies, and drinking. These activities potentially lead to a reduction in attention, awareness, decision making and performance. The diversion of attention from driving could also be influenced by in-vehicle technologies and external factors, such as external events and other road users [64].

- Affective states — represent the underlying experience of feeling, emotion or mood. These states could affect the response of a driver. The main types of affective states identified in the literature that could impact drivers' responses are: fatigue and emotional valence.
  - Driver fatigue — driver fatigue is defined as "the subjective feelings of tiredness and drowsiness" [48, 49]. Fatigued driving results in physical or mental disorders, which leads to a decline in attention, reaction time, recall, and vigilance [65]. In addition, fatigue driving increases the likelihood of driving errors, violations, inattention to road signs and users, and retarded decision making. Major causes of driver fatigue are long periods of driving, work pressure, sleep deprivation, weather conditions and time of the day [66, 67, 68].
  - Driver emotional valence — represents the extent to which an emotion is positive or negative e.g., happy, neutral, sad, angry, disgusted, frustrated, and stressed [69]. The driver emotion is influenced by external factors, such as activities in their personal lives, time pressure in commercial driving, traffic conditions and behaviours of other road users [70].
- Driving styles — are defined in terms of the manner by which the driver operates vehicle controls [71], such as how the driver accelerates, brakes, steers, changes gear etc. They are commonly categorised into two main classes: 'calm' driving style and 'aggressive' driving style. Unlike drivers' personality traits e.g. impatient and competitive, which are less likely to change over time, driving styles are dynamic depending on the driving environment and context [72, 73].

Several technologies have been adopted to capture data regarding driving behaviours, such as telematics, smartphones, video cameras, microphones, Electroencephalography (EEG), Functional Near-Infrared Spectroscopy (fNIRS), ElectroOculoGram (EOG), Sphygmomanometer (Sphygs), Electrocardiogram (ECG), Electromyography (EMG) and Galvanometer (GALV). The data produced by these technologies or sensors are analysed by researchers to help understand and characterise driving behaviours. Table 2.1 categorises the sensors identified in the literature to capture driver data according to the different types of driving behaviours. It can be observed that multiple physiological sensors i.e. i.e. EEG, fNIRS, EOG, Sphygs, ECG, EMG and GALV have been utilised to understand the relationship between physiological states and driving behaviours in controlled experimental settings. Incorporating these technologies in real-world driving environments is still impractical as it will require the adoption of related wearable sensors, which could be intrusive in terms of ethics and data privacy. As this research aims at developing an online driving risk assessment system, it focuses on analysing data generated by currently adopted in-vehicle technologies, such as telematics and cameras. In addition, due to the lack of data that captures the interaction and impact of driving behaviours and external factors on road safety, expert inputs are also considered.

Table 2.1: A review of the technologies and sensors that capture driver data for understanding and characterising driving behaviours

Driving behaviours	**Technologies/Sensors**
Driving styles	EEG [74] Telematics [14, 75, 76, 77, 78, 79] Smartphones [80, 81, 82, 83, 84]
Driver Emotion	ECG [85, 86] EEG [87] GALV [88, 89, 90, 91] Sphygs [92] EMG [88, 89, 90] Cameras [93, 94, 95, 96, 97] Microphone [98, 99]
Driver fatigue	EEG [100, 101, 102, 103] Microphone [104] fNIRS [105] Cameras [106, 107, 108, 109, 110, 111] ECG [101, 102] Telematics [112]
Driver distraction	EEG [87, 100, 113] Smartphone [114] fNIRS [7, 105, 113, 115] ElectroOculoGram (EOG) [116] Cameras [42, 43, 44, 45, 46, 117, 118]

2.3 Driver Characterisation

Automatic detection of driving behaviours from driver data plays a key role in the online assessment of driving risk because proper analysis of driving risk cannot be achieved without accurate and reliable information about the detected behaviours that impact road safety.

This section reviews data-centric AI methods employed in the literature to automatically detect driving behaviours from driver data. The purpose of this review is to present the opportunities and gaps in existing AI methods for driving behaviour characterisation. Due to the large number of data-centric AI methods employed for driver characterisation, the section is structured according to the following categories: 1) unsupervised learning methods; 2) conventional supervised learning methods; 3) deep learning methods; and 4) fuzzy systems.

2.3.1 Unsupervised learning

Unsupervised learning uncovers patterns or data representations from unlabelled data [119]. This is particularly applied to clustering and auto-encoders.

Clustering is an unsupervised learning technique in which objects are assigned to labels (clusters) based on their similarity or distance from other objects i.e., similar or nearby objects are found in the same cluster and dissimilar objects found in different clusters [120]. The clustering process minimises the distance between objects in a cluster and maximises the distance between different clusters. It is often used to identify patterns in unlabelled data. The main challenges in clustering are determining the right number of clusters and selecting the best clustering technique for the dataset, as there are numerous clustering techniques with different criteria for grouping data points (i.e., similarity measures). Several cluster number validity methods (cluster validity indices) have been proposed to obtain ideal number of clusters, such as Calinski and Harabasz [121], Hartigan [122], Silhouette [123], Davies-Bouldin [123], the Elbow method [124], partition coefficient and entropy [125]. These indices either try to maximise the separation between clusters or minimise the distance between points within clusters thereby making clusters more compact. In addition, ensembles of diverse clustering techniques are employed to overcome the challenge of 'selecting a particular clustering technique' and elucidate more robust clusters [126, 127].

In driving behaviour analysis, clustering techniques have been explored to recognise different driving styles based on the manner by which drivers operate vehicle controls. For example, k-means [128, 129, 130, 131], hierarchical clustering [57, 128], self-organised maps [132], and consensus of clustering methods [14] have all been employed to identify driving patterns based on driving data collected by sensors connected to the vehicle's controller area network (i.e. telematics) or smartphone sensors. The techniques aim at identifying groups of drivers with similar driving patterns emerging from the data collected. Subsequently, the driving features of the different groups of drivers (i.e., clusters) are analysed based on their distributions to identify the driving styles: 'calm', 'cautious', 'slow', 'normal', 'moderate', 'neutral', 'reckless', 'fast', and 'aggressive'.

Another recently explored unsupervised learning technique for driving style characterisation are auto-encoders. In auto-encoders, low-dimensional data (also known as latent features) are extracted from high-dimensional or unstructured data in an unsupervised manner, making it easier to discover underlying patterns within the data [133]. The main challenges in auto-encoders are deciding on the optimal number of features that produce reliable patterns and interpreting the non-linear extracted features. Auto encoders have been employed for driving style characterisation [132, 133, 134, 135]. For example, Siami *et al.* [135] employed deep auto-encoders to extract latent features from large smartphone trajectory data collected from 2500 drivers over 500,000 journeys. The data consisted of location, velocity, acceleration and change of direction information. The extracted features were fed to clustering techniques to identify six driving styles, namely warm stopping, cornering with medium speed, driving at normal speed, swerving at medium speed, weaving at high speed and cornering at high speed. Recently, Bandyopadhyay *et al.* [133] also employed a sequence-to-sequence auto-encoder on time series driver data made up of long-short term memory networks (LSTMs). The data was obtained from six drivers using smartphone sensor and consisted of inertial measurements and GPS. The extracted features were fed to a hierarchical clustering approach to identify three driving styles: normal, aggressive and drowsy.

The main remark after reviewing unsupervised learning approaches for driver characterisation is that the main source of data to recognise driving styles is vehicle control data, which can be obtained from telematics or smartphone sensors. However, the data is usually unlabelled due to its large volumes, which makes manual labelling very expensive and time consuming. Another finding from the review is that an ensemble of clustering techniques increases the degree of confidence and stability in the driving styles formed by merging the results of multiple clustering algorithms [14]. However, matching driving styles between clustering algorithms is not straightforward as different algorithms may generate different numbers of driving styles and moreover the optimal number of driving styles may be unknown. If the number of clusters specified is too small, the driving styles obtained may not effectively capture the different driving patterns in the data, while if the number of clusters is too large, the driving styles obtained may not represent any meaningful driving patterns. Therefore, obtaining the optimal number of driving styles using clustering techniques is important. Also, this approach may result in a number of drivers or driving patterns remaining unclustered. Furthermore, large longitudinal driving data have shown to reveal more underlying driving styles [135]. The aforementioned studies that explored unsupervised approaches however have limitations: 1) they do not consider environmental and external factors in their analysis, such as weather conditions, traffic conditions, time of the day, which have been illustrated by the TPB (reviewd in Section 2.2.1) to influence driving behaviours; 2) they do not consider the impact of other driving behaviours, such as driver emotions and distractions, which can be captured using images of drivers; 3) difficulty to provide meaning to the clusters identified; and 4) they do not provide relationships between driving styles and driving features, which is important for drivers to understand how to improve their driving styles, and decision makers to make informed decisions.

In this thesis, we explore intelligent methods to propose a multifaceted framework to driver characterisation using telematics and camera footage data. The methods developed in the framework tackle most of the aforementioned limitations and provide a holistic view of driver behaviour.

2.3.2 Supervised learning

Traditional supervised learning techniques, such as Logistic Regression (LR), Decision Trees (DTs), and Support Vector Machines (SVMs), have shown promising results in driver behaviour characterisation [45, 118, 136, 137, 138, 139, 140]. The techniques aim at detecting underlying patterns and relationships between the input data and the output labels. In driving behaviour analysis, the input data consist of driving features that represent the manner by which the driver operates the vehicle e.g. number of driving incidents, acceleration, speed etc, unstructured driver data e.g. images, audio, signals etc, or features extracted from unstructured driver data e.g. handcrafted features. While the output labels could be different categories of driving behaviours e.g. ‘tired’, ‘calm’, ‘rested’, ‘reckless’, ‘distracted’, ‘sad’ etc. The output labels can also be continuous values describing the extent of particular behaviours e.g., imagine driver distraction represented on a continuous scale from 0 to 100, where 0 represents ‘very attentive’ and 100 represents ‘very distracted’.

A review of some of the frequently employed supervised learning methods in driver behaviour characterisation is presented below:

1. Logistic Regression: LR uses a sigmoid function to find the linear dependence between the inputs and outputs [141]. The input features are combined linearly using weights to predict the outputs. The sign and size of these weights represent the direction and magnitude of the relationship between the features and the outputs. LR can be summarised using the following equation:

$$P(x) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n)}} \tag{2.1}$$

   Where *P(x)* is the expected probability of the output, n is the number of input features in the data, and b_n is the coefficient for the input feature X_n, which is learned from the data.

   LR in driver behaviour characterisation has been largely applied to vehicle control data to understand the relationship between the frequency of driving features or incidents and driving styles [142, 143]. It has also been applied to handcrafted or geometric features extracted from driver facial images and body posture images to recognise drowsy driving [142] and distracted driving [136, 137], respectively. The main advantage of using this technique is that it is easy to implement and interpret as the magnitudes of model coefficients are indicators of feature importance and their signs are indicators of direction of association. However, it is difficult to detect complex, non-linear relationships as it constructs linear boundaries.

2. Support Vector Machines: SVMs use hyperplanes to find the optimal decision boundaries that separate classes in a higher dimensional space [144]. The equation below shows a hyperplane (Equation 2.2) that classifies data objects as +1 when the objects lie above or on a hyperplane, and as -1 when they are below a hyperplane:

$$w.x + b = 0 \tag{2.2}$$

$$f(x) = \begin{cases} +1, & \text{if } w.x + b \geq 0 \\ -1, & \text{otherwise} \end{cases} \tag{2.3}$$

   where w is the set of feature coefficients, x is a feature set and b is the bias or intercept. Principally, SVMs try to find the optimal hyperplane that can separate data objects with minimum error.

   Similar to LRs, studies demonstrated SVMs outperform several conventional supervision learning methods in driver behaviour characterisation problems [118, 140, 145, 146, 147]. For instance, Chen *et al.* [147] reported better performance of SVMs in detecting normal and reckless driving behaviours compared to naïve bayesian classifier, k-nearest neighbor, and DTs using telematics data. Liang *et al.* [145] reported better performance of SVMs in recognising driver distraction state compared to LR using eye movement and telematics data. However, SVMs are difficult to interpret compared to LR and DTs.

3. Decision Trees: Another commonly applied supervised learning approach in driver behaviour characterisation is DTs, mainly due to their capability in capturing non-linear relationships, and the ease in understanding and interpreting their results. They are also easy to implement as the require little data preparation and can easily handle numerical and categorical data. They are hierarchical models that contain decision nodes and edges [148]. The nodes represent the conditions for splitting the data or the final outcomes, while the edges represent the decisions. The algorithm to define the structure of DTs is as follows:

   (a) The algorithm begins with the entire feature set, S.
   (b) On each iteration, it goes through all the features in S and calculates the information gain of the features.
   (c) The feature with the largest information gain is selected as the root node and removed from S.
   (d) The tree is then split by the selected feature according to the values that produced the largest information gain.
   (e) The algorithm goes back to step (b) to determine the next node, until S is empty.

   The visual representation of DTs makes them useful decision making and reasoning tools for driver behaviour analysis [45, 138, 139]. Ensembles of DTs have been also employed in driving behaviour, such as Gradient Boosting [149], Adaptive Booster [150], Bagging [151] and Random Forests (RFs) [152].

4. Bayesian learning: Bayesian networks (BNs) also known as 'belief networks' provide a simple way of applying Bayes Theorem to complex problems. BNs are probabilistic graphical models for representing conditional dependencies between a set of variables using some conditional independence assumptions [153]. They aim at updating the probability for a hypothesis as more evidence or information becomes available. The graphical representation of the models as nodes and edges makes them interpretable and easy to understand. The nodes represent the variables, such as driving features, while the edges that connect the nodes indicate the relationships between the variables. A variable can be indirectly influenced by another variable via edges through other variables, and there should be no graph paths starting and ending at the same node (no directed cycles). The network captures the joint probabilities of the variables represented by the design of the graph.

   Figure 2.3 illustrates a simple BN for predicting drivers' fatigue. The network has three nodes; 'Driving in day or night', 'Number of hours rested', and 'Fatigue'. 'Driving in day or night' and 'Number of hours rested' are the independent variables, while 'Fatigue' is the dependent variable. Probabilities of independent variables could be used to compute the probability of the dependent variable. The probabilities and structure of BNs are obtained from data or domain experts. For example, Yan *et al.* [154] utilised BNs to extract important features affecting driving using data from a driving simulator, while Rashwan [155] used expert opinion to develop a conditional distribution table for predicting driver fatigue. BNs have been employed in the prediction of aggressive driving styles [156], fatigue [155, 157] and driver distraction [158, 159]. In Han *et al.* [156], the authors developed a system predict the probability of a driver being normal or aggressive driving style using a BN. However, only two driving features were analysed i.e., vehicle speed and throttle opening. In addition, Al-Sultan *et al.* [157] demonstrates BN's ability to incorporate contextual information into the characterisation of driving behaviour. However, it is difficult to design BNs, estimate their conditional probabilities and interpret their complex relationships for datasets with many features. Most importantly, BNs do not handle imprecision and subjectivity in information, which is inevitable in the description of driver behaviour [160].

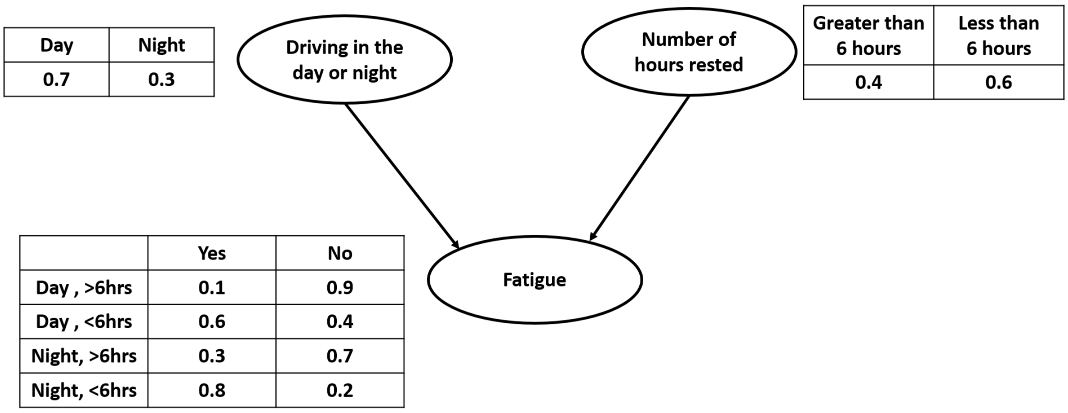


Figure 2.3: A simple Bayesian network to predict drivers' fatigue with conditional probability tables for the different variables.

2.3.3 Deep learning

Deep learning approaches outperform traditional machine learning when processing unstructured data for driver characterisation. DNNs consist of densely interconnected processing elements (i.e. neurons), which capture non-linear relationships between input features and outputs. In driver behaviour characterisation, such relationships are found in unstructured, complex data e.g., images, sound, and physiological signals. These unstructured data are mainly explored to recognise observable driving behaviours. The main categories of DNNs to recognise driving behaviours are: Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs).

1. Convolutional Neural Networks: CNNs [161, 162, 163, 164] are neural networks consisting of filtering (or convolution), pooling and activation layers. The inputs go through the convolution layer, where they are filtered to produce stacked smaller dimensional features (i.e., feature maps), which capture various spatial patterns in an image such as edges, shapes, and intensity in pixels. The stacked feature maps go through a pooling layer, which summarises their

representations using a sample-based discretisation process, such as max pooling [165] that calculates the maximum value of different patches in the feature maps. The pooling process helps in reducing the number of trainable parameters and model complexity. The activation layer later converts the stacked downsampled data into more computationally efficient features depending on the activation function used e.g. Rectified Linear Unit (ReLU) activation function [166] converts all negative values to zero and maintains all positive values, thereby, speeding up computation of derivatives. These filtering, pooling and activation layers allow CNNs to learn hierarchical discriminative features. Figure 2.4 presents an example of the convolution operation on a 2D image where zeros are added to its boundaries (i.e., padded with zeros) to enable uniform analysis of the image. The convolution operation uses a kernel of size 3*3, a 2*2 pooling patch and a ReLU activation function. The filters slide through the images and perform mathematical computations on the pixels of the images. The sliding operation is controlled by the value of the stride e.g. a stride of 2 means the kernel slides by 2 columns sideways and 2 rows downwards indicated by the red arrows in the diagram.

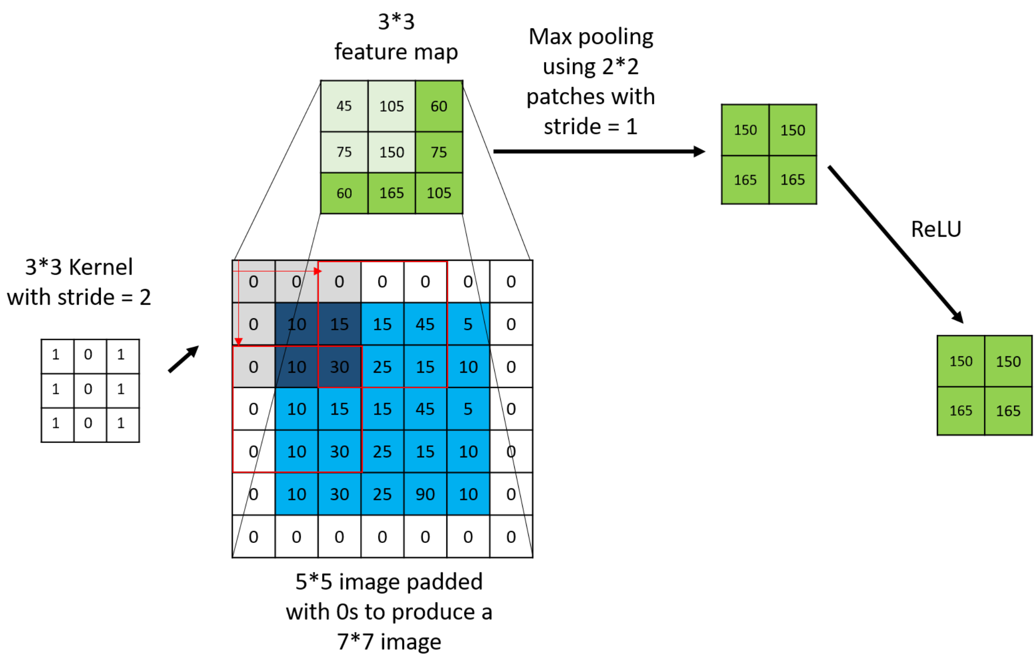


Figure 2.4: An example of the convolution operation on 2D image using a kernel of size 3*3

In driver behaviour characterisation, several studies have applied CNNs to unstructured driver data, such as driver images [167], audio [168] and physiological signals [169, 170] to detect driver distraction [170] and affective states [167, 168, 169]. Others perform transfer learning from the ImageNet dataset [171] to take advantage of its large and diverse images in initialising model weights [172, 173, 174, 175]. Existing CNN architectures have demonstrated good performance in automatically recognising driving behaviours, such as VGG [176], Inception [177], ResNet [178] and DenseNet [179]. These deep learning architectures differ in the number of convolution layers, connections between layers and model features (e.g., activation functions), and they were introduced to solve different issues in training and performance of CNNs. For example, Inception was introduced in 2014 to deal with the uncertainty in choosing the kernel size of convolutional layers by using multiple kernel sizes in each convolutional layer, while ResNet was developed in 2015 to handle the vanishing gradient problem by introducing skip connections between layers [180]. A limitation of these architectures is that they fail to capture temporal dynamics of behaviours. For example, if the temporal context is not considered when detecting a driver's distraction state, a single image frame selected from a sequence of frames representing the driver's hand movement from the steering wheel to the gearshift may be mistaken for reaching to the phone or radio. Such temporal changes in driving behaviours have led to the application of recurrent neural networks (RNNs), hybrids of CNNs and RNNs and stacked RNNs in driver behaviour analysis.

2. Recurrent Neural Networks: RNNs are deep neural networks with feedback loops connecting the output of the previous state to the current state [181] as shown in Figure 2.5. This enables the network to remember what the model learned from previous time steps. Unlike simple CNNs that map a single input to an output, RNNs map a sequence of inputs to an output or a sequence of outputs.

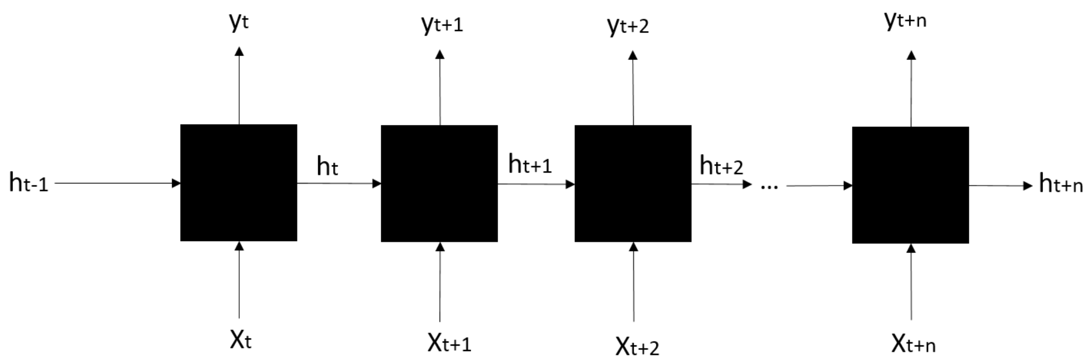


Figure 2.5: An illustration of feedback loops of a recurrent neural network, where ht is the hidden state at time t, X_t is the input at time t, and y_t is the output at time t.

The following equations represent the memory cell of a RNN, which consists of hidden states to store information from previous and current time steps:

$$h_t = f_h(W_h X_t + U_h h_{t-1} + b_h) \tag{2.4}$$
$$y_t = f_o(W_o h_t + b_o) \tag{2.5}$$

Where X_t is the current input data at time t, h_t is the current hidden state obtained from the current input data and the previous hidden state (h_{t-1})

and y_t represents the output. W, U, and b represent the weight matrices and bias vectors which need to be learned during training, while f_h and f_o are the non-linear activation functions.

Long Short Term Memories (LSTMs) [182] are an extension of RNNs capable of remembering information in longer sequences i.e., learning longer dependencies in the data. They consists of additional gates in the memory cell, namely: forget, input, and output gate layers. The input gate controls which state is updated. The forget gate controls how much information needs to be retained or forgotten, and the output gate decides which part of the cell state is outputted to the next LSTM unit. Figure 2.6 represents a simple LSTM memory cell to demonstrate the flow of information between the gates. The inputs to the LSTM cell are the current input data (X_t), the cell state of the previous time step (C_{t-1}) and the hidden state of previous time step (h_{t-1}).

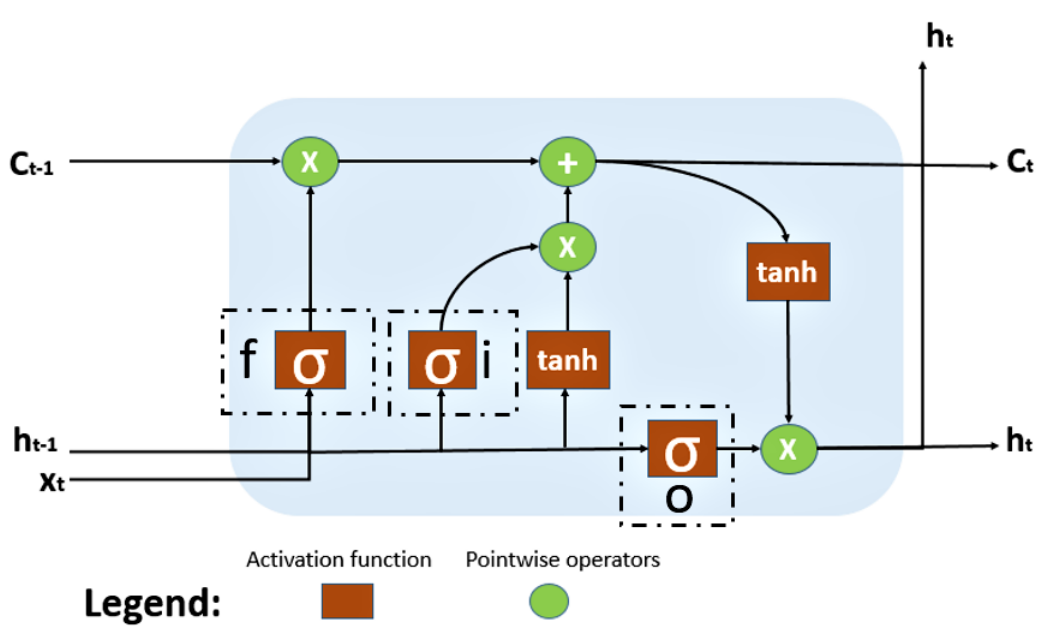


Figure 2.6: An LSTM memory cell with forget (f), input (i) and output (o) gates as illustrated in [2].

The interactions between the gates in the LSTM memory cell is given by the following equations

$$f_t = \mathrm{sigm}(W_f X_t + U_f h_{t-1} + b_f) \tag{2.6}$$
$$i_t = \mathrm{sigm}(W_i X_t + U_i h_{t-1} + b_i) \tag{2.7}$$
$$o_t = \mathrm{sigm}(W_o X_t + U_o h_{t-1} + b_o) \tag{2.8}$$
$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c X_t + U_c h_{t-1} + b_c) \tag{2.9}$$
$$h_t = o_t \odot \tanh(C_t) \tag{2.10}$$

Where f_t is the forget gate's activation vector, i_t is the input or update gate's activation vector, and o_t is the output gate's activation vector. W, U, and b

represent the weight matrices and bias vectors which need to be learned during training.

Other variations of LSTMs have been introduced to improve sequence learning tasks, such as stacked LSTMs [183], Gated Recurrent Units (GRUs) [184] and bidirectional networks [185].

RNNs and their variations have been explored for recognising driver distraction postures using telematics data [186, 187, 188]. However, for driver distraction recognition, analysing driving posture images using CNNs has shown to produce better results compared to using telematics data [42, 43, 170, 172, 186]. In addition, with the availability of publicly available driving posture datasets, this thesis addresses the limitations of CNNs in learning the temporal dynamics of driving postures to improve the recognition of driver distraction by exploring the combined benefits of CNNs and RNNs.

Furthermore, facial images have been the main data source for characterising driver affective states i.e. driver emotion and fatigue as researchers have investigated different variations of CNNs and RNNs using facial images to distinguish between different affective states [55, 109, 189, 190, 191, 192]. However, the results reported by these studies are still far from deployment in terms of performance, especially for the purpose of road safety. For instance, the best performance reported using the most comprehensive publicly available affect database is 0.620 [193, 194, 195, 196, 197], where 1.0 is perfect agreement between actual and predicted values. Therefore, there is still much room for significant improvement to attain remarkable performance. Furthermore, compared to other data sources that can be easily anonymised to protect users' privacy e.g. telematics, driving posture images, facial images are at risk of data privacy as the identities of people used to train affective models can be exposed in the process. Therefore, in addition to exploring DNNs to improve driver affect recognition, this thesis examines effective techniques to preserve the privacy of data subjects, such as facial action units that anonymise facial data while maintaining the usability of the dataset [196, 197, 198, 199], and federated learning that processes users' facial images in their local machines and only send their locally trained models back to the developer's machine for augmenting the final model [200, 201].

To conclude, it is worth mentioning some challenges of DNNs: 1) model complexity [202, 203]; and 2) model interpretability [204, 205]. Most DNNs consist of millions of parameters which need to be optimised and several hyper-parameters which need to be set to control the learning process e.g., number of neurons, number of layers, learning rate, filter size, and batch size. Tuning these hyper-parameters can be stochastic and time consuming. In addition, the model complexity of DNNs makes their decisions and processing difficult to understand and explain to non-ML experts.

2.3.4 Fuzzy systems

The identification and description of driving behaviours are uncertain, imprecise and subjective. Trying to precisely characterise driving behaviours into specific categories, such as 'calm', 'normal', 'aggressive', 'sad', 'happy', 'angry', is difficult due to different subjective interpretations of behaviours, context and imprecision of intelligent driver characterisation systems. Even when introducing contextual information as proposed by TPB and RHT in Section 2.2.1, the precise effect of contextual factors is difficult to determine as the relationship between the factors and driving behaviours is not clear-cut due to their interactive effects. Therefore, to effectively characterise driver behaviour and incorporate contextual information in the assessment of driving risk, there is need for a suitable computational approach that can handle uncertainties and imprecision in the information about driving behaviours, contextual factors and their relationships.

Fuzzy Logic Systems (FLSs) have shown their effectiveness in handling those uncertainties for a wide range of applications, such as, medicine [206, 207], insurance [208], finance [209], environmental planning [210] engineering [211] and cyber security [212, 213]. FLSs model and represent input and output uncertainties (e.g., uncertainty associated with subjectivity, uncertainty associated with words, vagueness in definition) using input and output fuzzy sets, respectively [18]. The systems consist of three main stages: 1) Fuzzification , 2) Inference, and 3) Defuzzification.

Before the above stages of a FLS are described, we first define a fuzzy set (FS) which is the fundamental element of a FLS. In a set theory, an object x either belongs, or not, to a set A. Whereas in fuzzy set theory, an object x can belong to many sets with different degrees of membership. That is, a fuzzy set A is a pair (U, μ) where U is the set domain and μ is the membership function. For an element $x \in U$,

$$\mu_A(x) = f(x) \tag{2.11}$$

where $f(x)$ is a membership function that maps x to a membership degree in [0, 1].

In fuzzification, a crisp input (e.g. driving feature) is converted into membership degrees using singleton and non-singleton FSs. Commonly used non-singleton FSs in real world applications are triangular, Gaussian, trapezoidal, and sigmoidal membership functions [214]. We describe some of these membership functions below:

- **Triangular membership functions:** A triangular membership function has three key points (l, m and u): the lower limit (l), the expected value (m) and the upper limit (u). A triangular membership function is represented by the following equation:

$$\mu_F(x; l, m, u) = \begin{cases} \frac{x-l}{m-l}, & l < x <= m \\ \frac{u-x}{u-m}, & m < x <= u \\ 0, & otherwise \end{cases} \tag{2.12}$$

- **Trapezoidal membership functions:** Trapezoidal membership function: A trapezoidal membership function consists of four points (l, m, n and u): the lower limit (l), the range for the expected value [m,n], and the upper limit (u). A trapezoidal membership function is represented with the following equation:

$$\mu_F(x; l, m, n, u) = \begin{cases} \frac{x-l}{m-l}, & l < x < m \\ 1, & m <= x <= n \\ \frac{u-x}{u-n}, & n < x < u \\ 0, & otherwise \end{cases} \tag{2.13}$$

- **Gaussian membership functions:** The Gaussian membership function consists of the mean and standard deviation of the data (m,σ). The 'bell' shape is defined as follows:

$$\mu_F(x; m, \sigma) = e^{\frac{-(x-m)^2}{2\sigma^2}} \tag{2.14}$$

- **Sigmoidal membership functions:** This membership function is represented by the sigmoid function:

$$\mu_F(x; a, b) = \frac{1}{1 + e^{-a(x-b)}} \tag{2.15}$$

  Where a is the center value of the sigmoid and b is controls the 'width' of the sigmoidal region.

Fuzzy inference maps crisp inputs to fuzzy sets. Subsequently, they process and aggregate the input fuzzy sets using fuzzy rules and operators to produce output fuzzy sets e.g., Mamdani rule-based inference method [215]. Below is an example of a fuzzy rule represented using an IF/THEN statement:

IF ***driver's speed*** *is* ***'high'*** *AND* ***road*** *is* ***'clear'*** *THEN* ***driving style*** *is* ***'normal'***

The final stage in a FLS is defuzzification. Defuzzification is the process of transforming the output fuzzy sets into a single real value or crisp output [216]. Even though inference systems model uncertainties using fuzzy sets, the outputs of FLSs are usually crisp values i.e., defuzzified output fuzzy sets. This is because FLSs are most often used in control systems, which perform certain actions when the system outputs a particular value. For instance, a driver-assistance system could provide feedback to a driver to adjust their driving style or even automatically initiate vehicle brakes when their driving style reaches a certain value. There are several methods of defuzzification such as Center of Gravity (COG), Mean of Maxima (MOM), Center of Mean (COM) and Midpoint of Area (MOA).

FLSs have been examined in driving behaviour characterisation to capture uncertainties and imprecision in driver features [217, 218, 219, 220, 221]. For instance, Aljaafreh *et al.* [217] employed FLS to model and capture uncertainties in two driving features i.e., acceleration and speed, for the prediction of four driving styles i.e., below normal, normal, aggressive and very aggressive driving. The authors did not mention the size of the dataset nor how the data is collected. Their membership functions were obtained from real data and domain experts. Similarly, Imkamon *et al.* [218] used FLS to predict a driver's aggressive level between 1 and 3 (where

1 is a normal driver and 3 is an aggressive driver) by modelling uncertainties in multimodal information from three sensors, an engine control unit (ECU) reader, an accelerometer and a camera. The authors used data obtained from a single driver during 50 minutes of driving and subjective evaluation from three passengers to test their system. The results showed about 70% similarity with the opinions of the passengers. In a different study that focused on driver fatigue, Bergasa *et al.* [221] explored FLS to detect the level of driver fatigue by combining driver facial features extracted from images, such as percentage of eye closure, eye closure duration, blink frequency, nodding frequency, face position, and fixed gaze. The authors utilised data from 10 drivers using a driving simulator. The performance of their system was measured by comparing the performance to results obtained by manually analysing the data on a frame-by-frame basis. They found that their system's performance decreases during daytime and with drivers wearing glasses. The main limitations of the above studies are the size, quality and diversity of their data (less than 10 drivers). In addition, limited work has been done with respect to HGV drivers.

Furthermore, fuzzy sets provide meaningful representations of inputs in terms of linguistic terms, and fuzzy rules provide valuable insights about the mappings between inputs and outputs. These characteristics make FLSs inherently understandable by humans, and an easy and effective method to fuse heterogeneous information [222, 223, 224].

2.3.5 Summary of driver characterisation

The review of intelligent methods employed for driver characterisation is summarised in Table 2.2. As revealed by the psychological theories of driver behaviour in Section 2.2, the real-world driving environment is at the confluence of the different driving behaviours i.e., driving styles, driving postures and affective states. However, we observe in the literature a general lack of intelligent approaches to simultaneously characterise these traits for a holistic view of drivers' actions, feelings and behaviours during driving.

For stand-alone analysis of the different facets of driver behaviour, we found out that unsupervised learning methods are generally explored for elucidating driving styles using telematics and smartphone sensor data due to the difficulty in labelling the data. However, the sensor readings have uncertainties, and the driving styles are dependent on the data. Therefore, there is need to explore intelligent methods that can handle data uncertainty and imprecision in interpretations. In addition, more diverse and comprehensive data (i.e., data captured from a large cohort of drivers completing several journeys throughout the year in multiple road types and weather conditions) is needed to identify reliable and core driving styles. The review also revealed that conventional supervised learning models are most efficient for labelled structured data due to their low computational costs compared to deep learning approaches and interpretable results. However, structured driver data, such as vehicle operation data collected using telematics and smartphone sensors are usually difficult and expensive to label due to their large data size. Therefore, unsupervised and semi-supervised methods could be utilised to label large driving data (thousands of data instances) and supervised techniques employed to train models for predicting the driving styles of new driving patterns.

For driving behaviours that can be best recognised using unstructured driver data (e.g., images) i.e., driver distraction and affective states, DNNs are most suitable to process the data. CNNs automatically extract static discriminative features from large and diverse datasets of images, while RNNs capture temporal dynamics in sequences of data or among image pixels. However, the main drawback of DNNs is the difficulty to interpret their predictions. Furthermore, as earlier mentioned in Section 2.3.3, anonymising facial images is more difficult compared to other driver data sources as conventional data anonymisation techniques such as blurring and pixelation could make the images no longer useful for the task. Therefore, processing facial images requires privacy-preserving strategies to protect drivers' identities and sensitive information, while ensuring the usability of the data.

Lastly, the review suggests FLS as a possible solution to model and capture uncertainties in driving data injected by sensor readings, human subjectivity, and imprecision in the description of driving behaviours and external factors. In addition, it is an effective information fusion strategy for combining the heterogeneous characterisations of driving behaviours and external contextual information. However, designing complex FLSs with several variables requires hierarchical fuzzy controllers [225, 226] to preserve interpretability and prevent rule-explosion problem (i.e., the number of rules in the system increases exponentially with the number of variables involved).

2.4 Driving Risk Assessment

After proper and automatic characterisation of driving behaviours by intelligent driver assessment systems, the decisions are sent as inputs to a driving risk assessment system to assess their impact on road safety. This section reviews the literature on driving risk assessment, which consist of post-hoc analysis of driving risk and online data-driven driving risk assessment. Post-hoc analysis of driving risk consists of statistical analysis of driving data captured from experiments, surveys or real-world settings. It is used to draw insights, conclusions or determine the effects of different factors on road safety. While online data-driven driving risk assessment is the automatic processing of data streams about drivers' actions and states, vehicle characteristics and environmental conditions to produce decisions for the prevention of road incidents or accidents.

First, this section reviews studies that employ post-hoc analysis on data from questionnaires, driving simulators, physiological sensors, and real-world settings to understand the impact of driving behaviours and external conditions on driving risk. Popular statistical techniques employed are Path analysis [27], p-value of means [227], Logistic regression [227, 228], Poisson regression [11], and Pearson correlation analysis [7, 228]. Subsequently, the literature on data-driven intelligent driving risk assessment is reviewed.

Table 2.2: A summary of the review of intelligent methods applied to understand and characterise driving behaviours.

Methods	**Category**	**Driving behaviours**	**Data Types**	**Strengths**	**Limitations /Challenges**
Clustering	Unsupervised	Driving styles	Structured	-Uncover driving patterns -Labelling of driving data	-Difficult to interpret -Do not capture uncertainties -Do not consider context
Representation learning	Unsupervised	Driving styles	Structured and unstructured	-Dimensionality reduction	-Difficult to interpret -Do not capture uncertainties -Do not consider context -Require a large dataset
Logistic Regression	Supervised	All behaviours	Structured	-Easy to implement -Easy to interpret	-Difficult to detect complex relationships -Do not capture uncertainties -Inefficient for high dimensional data
Support Vector Machines	Supervised	All behaviours	Structured	-Efficient for high dimensional data	-Difficult to interpret -Do not capture uncertainties -Inefficient for large data
Decision Trees	Supervised	All behaviours	Structured	-Efficient for high dimensional data -Easy to interpret	-Difficult to interpret for more complex tree algorithms e.g., Extra Trees -Do not capture uncertainties
Bayesian learning	Supervised	All driving behaviours	Structured	-Easy to interpret -Capture Uncertainties	-Difficult to build complex systems
Convolutional neural networks	Deep learning	-Distraction -Affective states	Unstructured	-Efficient for static data	-Difficult to interpret -Do not capture temporal dynamics -Computationally expensive
Recurrent neural networks	Deep learning	All behaviours	Structured	-Efficient for temporal data	-Difficult to interpret
Fuzzy systems	Linguistic	All behaviours	Structured	-Efficient for combining multimodal information -Easy to interpret -Capture uncertainties	-Difficult to build complex systems with many variables -Difficult to evaluate performance

2.4.1 Post-hoc analysis of driving risk

Historically, driving errors (e.g., harsh braking), violations (e.g., wrongful overtaking and over speeding) and driver mental states (e.g., fatigue and emotions) have been identified as the leading determinants of road safety [5, 71, 229]. As a result, extensive work has been done to assess the impact of these risky driving behaviours on road safety and the factors influencing the behaviours. The main post-hoc method employed to assess driving risks is the analysis of data collected from questionnaires and surveys [7, 12, 49]. These studies distribute questionnaires to drivers or managers asking them to provide their opinions about the effects of different driving behaviours, external factors and company policies on road safety, and how often the drivers engage in risky behaviours. For instance, in an earlier research on how drivers' affective states affect driving risk [6], the authors found a strong association between driving violations and occasions of anger a driver experiences while driving. Their study recruited 104 non-commercial drivers in Britain, and provided questionnaires to capture the intensity of drivers' anger and violations. Correlation analysis was used to identify the relationship between anger and driving violations. In a different study engaging with bus managers in Taiwan [11], the authors utilised survey data about the effects of environmental (e.g., urban and intercity roads) and organisational factors (e.g., ratio of driver to non-driver staff and company capital) that affect road safety. The authors used Poisson regression to uncover the relationship between accidents and the factors investigated. Their results revealed higher risk of accidents in urban roads compared to rural and freeway, and lower risk of accidents for larger bus companies. Forty two companies were considered. Tseng *et al.* [12] analysed questionnaire data from HGV drivers in Taiwan about the leading factors that affect road safety. The authors collected data for 2,101 HGV drivers from a national survey in 2012. Their results reveal night driving and sleep quality as the main factors affecting road safety of HGV drivers. More recently, Hammad *et al.* [230] analysed questionnaire data from 50 drivers in Pakistan about the environmental factors leading to road traffic accidents. Their results showed that rainfall, severe coldness, fog, and heat conditions were directly related with the occurrence of road traffic accidents. In addition, Salmon *et al.* [231] analysed questionnaire data collected from 316 non-commercial drivers in Australia to understand the factors influencing drink and drug driving, distraction and inattention, speeding, fatigue, and failure to wear a seat belt. Their results identified the following system factors: road safety policy, transport system design, road rules and regulations, and societal issues.

Even though drivers have been the main participants of these studies because of their hands-on information about the contextual factors, other crucial stakeholders in the driving community also possess valuable and supplementary insights about the factors, such as road safety officers, transport managers, and traffic safety researchers. These stakeholders interact closely with drivers, policy makers and other road users to improve safety in road networks. Therefore, there is need to engage with these stakeholders to obtain more comprehensive and reliable insights about the influence of driving factors on road safety. In spite of the advantages of engaging with stakeholders, a major challenge is the processing of diverse information obtained from the stakeholders as insights obtained from humans may differ [232], and the information from the stakeholders need to be effectively aggregated. Fuzzy Expert Systems (FESs) [19] show great potential to tackle those challenges from a computational perspective as they can effectively capture and aggregate insights from humans using linguistic fuzzy IF-THEN rules, while addressing ambiguity in information using fuzzy sets.

Apart from surveys, data from driving simulators are also examined to obtain insights. Driving simulators are controlled experiments that closely simulate real-world driving conditions, including external influencers. They consist of sensors to capture vehicle control data, driver physiological signals, driver eye movements, facial images, and images of driver posture. Data collected from these controlled environments have been explored to assess the effects of contextual factors on drivers' performance [233, 234, 235, 236]. For example, Du *et al.* [237] investigated the effects of fatigue on drivers' performance (steering wheel movements, lane position and longitudinal speed) using driving simulation data from 24 participants. Using paired sample t-test, they found a significant difference in performance between fatigued and alert driving. While Hamdar *et al.* [235] examined the effects of weather conditions (i.e. foggy, icy and wet weather conditions) on drivers' performance (speed and deceleration). The authors used multivariate analysis of variance to analyse driving simulation data, and found that drivers' performance are affected by adverse weather conditions. Later, Li *et al.* [236] investigated the influence of traffic congestion on driver behaviour using driving performance measures, eye movement measures, and electroencephalogram (EEG) measures captured from a driving simulator. Their results showed more aggressive driving behaviour while driving in post-congestion situations.

Post-hoc analysis has also been employed on naturalistic data captured in real-world driving settings. For instance, Zou *et al.* [238] investigated the impact of climate and non-climate factors on fatal traffic accidents using a dataset of fatal traffic accident frequency in California and Arizona. Their results suggest that temperature and precipitation can significantly affect the frequency of fatal traffic accidents. Also, non-climate factors, such as rural roads and vehicle performance can significantly influence fatal traffic accidents.

2.4.2 Online driving risk assessment

This section reviews literature on driving risk assessment that use computational and artificial intelligence methods to process data and predict the driving risk level of driving scenarios. Such intelligent systems are important for providing proactive traffic intervention strategies, real-time optimisation and safety management of driver journeys, and timely and reliable support to drivers.

Due to the lack of ground-truth or labelled driving risk datasets, hybrid unsupervised learning methods and expert systems have been mostly employed to label driving patterns and classify them according to their level of risk. Such as, clustering [239, 240, 241], rule-based method [15], clustering + XGBoost [239, 240, 242], clustering + SVMs [239], clustering + Nearest Neighbors [243], FLS [244] and BNs [154]. The hybrid clustering methods group drivers with similar driving patterns using an unsupervised learning method, assign a risk level to each group according to the intensity of features in the groups using cluster labelling strategies [14, 245] and subsequently, employ a supervised learning method to learn the relationship between the features and the categories of driving risk. For example, Fernandez and Ito [244] presented a fuzzy rule-based system to assess the risk of driving styles (i.e., 'passive', 'normal', 'aggressive' and 'dangerous') using driver's age, percentage of times the driver uses the accelerator and brake pedals along the route, the percentage of times under the low speed limits, the percentage of times over the high speed limits, and the percentage of times driving within the right speed limits. Their results showed two rules that lead to 'passive' and 'aggressive' driving. In a different study that used driving simulation data [154], the authors employed Bayesian networks and logistic regression to establish a driving risk prediction model. Their results on a dataset from non-commercial drivers revealed driver experience as the most important driver feature in predicting driving risk.

More recently, Shi *et al.* [240] used Fuzzy C-means to uncover and label driving patterns according to their level of risk. Subsequently they employed extreme gradient boost (XGBoost) to learn the linkages between driving features and corresponding risk level, and automatically predict the risk of driving patterns. The authors explored driving features of non-commercial drivers from Next Generation Simulation vehicle trajectory database [246], and uncovered four driving risk categories: 'safe', 'low risk', 'medium risk' and 'high risk'. Shangguan *et al.* [239] explored fuzzy c-means and XGBoost to develop an online driving risk assessment system. The authors employed a rolling time window approach to model their data for real-time driving risk prediction. Their results on car-following events from Shanghai Naturalistic Driving Study show an optimal time window length of 0.5 seconds with a frequency of 0.1 seconds, and vehicle speed as the most risky driving feature. Mehdizadeh *et al.* [15] trained machine learning models to predict the likelihood of observing at least one safety critical event (close following, harsh braking, activation of rolling stability system, and activation of collision mitigation system) over a 30-minutes driving window. They utilised trajectory (e.g., time, GPS and speed) and safety critical events collected by a large U.S.-based trucking company from April 1, 2015 to March 31, 2016 for 496 truck drivers. The authors supplemented the data with weather information (e.g., precipitation and wind speed) and traffic flow information. Multiple supervised learning methods (e.g., nearest neighbours, ensemble of trees, SVMs) were employed to learn the relationship linking drivers' vehicle control, past occurrence of events, weather conditions, and traffic flow to the frequency of safety critical events over a 30-minutes driving window. Their results showed 'vehicle speed' and 'hour of the day' as the main predictors of driving incidents.

Deep neural networks have also been employed on big structured datasets to predict the risk of driving patterns [16, 247]. For instance, Bian *et al.* [247] proposed a driving risk assessment system based on deep neural networks. The system was applied to telematics data i.e., mileage, speed, engine load, engine temperature and fuel consumption, collected from 1347 non-commercial drivers in China within three months (August 2016 and November 2016). Their system classified driving styles into five risk levels where level 5 is the most risky. Their experimental results showed better performance compared to conventional supervised learning methods i.e., SVMs and RFs. Subsequently, Hu *et al.* [16] employed a semi-supervised deep learning method to assess driving risk. Their method consisted of LSTM encoder-decoder networks and CNNs to classify driving risk using telematics data from 21,124 vehicles in Germany. The vehicles consisted of cars and trucks. The authors manually annotated a small subset of the data (data from 1250 vehicles) with the following risk labels: 'low', 'normal', 'high', and 'extremely high'. The remaining unlabelled dataset was trained in an unsupervised approach using the LSTM encoder-decoder networks coupled with CNNs, and the small labelled dataset is used to fine-tune the performance of model using a supervised learning approach. Their results showed an accuracy of 95.82% when only 2.5% of the total data was labelled and used to fine tune the model.

2.4.3 Opportunities to improve driving risk assessment

This section presents gaps and opportunities identified in the literature of driving risk assessment that are addressed in this thesis. As mentioned in Chapter 1, the aims of this thesis are to develop an intelligent system that can automatically assess the impact of HGV driving in accordance with the characteristics of a real-world driving environment presented in Figure 2.2.

(a) Heavy goods vehicle driving

The majority of the studies in the literature that explore intelligent approaches to assess and predict driving risk focus on non-professional drivers i.e., people whose jobs are not driving. The driving behaviours of non-professional drivers have been shown to be different from those of professional drivers due to factors, such as driver training [248, 249], objectives [250], feedback systems [251], working conditions [252], duration of job shifts [229, 253, 254]. For example, Oz *et al.* [229] investigated the proneness to fatigue in professional and non-professional drivers. The authors recruited 234 male professional and non-professional drivers from four different driver groups (taxi drivers = 69; minibus drivers= 63; HGV drivers = 64; and non-professional drivers = 38) to report on the proneness to fatigue caused by the duration of their journeys. Their results show that HGV drivers are more prone to fatigue compared to non-professional drivers due to long and lonely journeys. Therefore, there is need for more work to be done in driving risk assessment with regards to professional and commercial drivers e.g., HGV driving.

(b) Imprecision in the definition of driving Behaviours

The description of driving behaviours are imprecise due to imperfect driver characterisation systems and data, as well as human subjectivity. That is, it is difficult for an individual or AI system to precisely define or predict driving behaviours. For example, it is difficult to provide a precise level of driver fatigue on a scale from 0 to 100 where 0 is 'alert' and 100 is 'fatigued' or driving style on a scale from 0 to 100 where 0 is 'slow' and 100 is 'aggressive'. The current literature on online driving risk assessment do not consider imprecision of driving behaviours, which could lead unreliable and unfair driving assessments. Therefore, assessing driving behaviours requires computational techniques that can handle imprecision in driving behaviours.

(c) Imprecision of AI models

The predictions produced by AI models are imprecise especially in the dynamic and complex HGV driving environment, which consists of numerous contextual factors, evolving traffic laws, constantly changing consumer behaviours, new technologies and infrastructure. These factors constantly influence drivers' actions and behaviours, making the decisions of the AI models (that have been trained on historical data) uncertain. Furthermore, additional uncertainty is introduced by the probabilistic outcomes of the AI models. For example, how do we differentiate between 60% and 65% probability outcomes of a driver's behaviour when assessing its impact on road safety? Once Again, FLSs are capable of modelling such uncertainties in AI model decisions using suitable FSs.

(d) Context-aware driving risk assessment

Another opportunity for improvement is the incorporation of contextual information into the assessment of driving. The few studies that explore intelligent approaches for online HGV driving risk assessment [15, 17, 241] do not consider the real dynamics of HGV driving (as described in Section 2.2.1), which consist of the occurrence and interaction of drivers' personal traits and external contextual factors e.g., driver's affective states, driving styles, attentiveness, weather conditions, road type and traffic conditions. This potentially produces incomplete assessments of the driving environment. Therefore, there is the need for multi-modal data streams that capture the interaction between the manner by which drivers operate their vehicles, driving postures, mental states and environmental conditions. However, due the lack of multi-modal data streams, this thesis (Chapters 4 and 5) engages with key stakeholders in the HGV driving sector to capture information about those interactions and embed the knowledge into the assessment of HGV driving. The hope is to achieve more context-aware, comprehensive and reliable assessments.

(e) Expert knowledge

An effective and reliable assessment of HGV driving is based on a nuanced understanding of the highly complex and dynamic environment, which is not available based on current data sources, but can at least be partially obtained from key stakeholders of the HGV driving sector as shown in the literature (Section 2.4.1). However, several challenges arise when capturing and modelling knowledge from a wide variety of stakeholders as responses obtained from humans may differ due to different levels of imprecision in perception, experiences and expectations [232, 255]. Thus, it is expected that different stakeholders —even though they have similar roles— provide different answers to questions due to ambiguity in the characterisation of driving behaviours, different levels of indecision and personal experiences [255]. For example, the precise level of driving risk when a driver is 'angry' and the weather is 'rainy' may be difficult to determine. In addition, stakeholders with different roles may have varying viewpoints due to their distinct responsibilities and expectations [256]. For example a group of researchers may have different opinions compared to a group of managers as managers may focus more on factors that optimise the delivery of goods and services in their companies. The possible differences in the responses provided by stakeholders must be effectively modelled to provide a comprehensive, reliable and clear representation of knowledge to support decision-making. In addition, there is the need to explore information fusion approaches and algorithms to embed the knowledge from stakeholders into the assessment of driving performance and risk. A modelling technique called the Interval Agreement Approach (IAA) [257] was developed to tackle the variability and imprecision of human insights by capturing agreement across sources of information provided by participants via surveys. Chapter 4 provides the rational for using this approach to model the variability and imprecision in the responses of stakeholders about the interaction and impact of HGV driving risk factors. The chapter will also provide an approach to embed the contextual information into the assessment of HGV driving performance and risk.

(f) Interpretability of assessments

Interpretability of driving risk assessments are important for verification, diagnostics, usability, and improvement [258]. Interpretability facilitates the understanding of how and why given assessments were produced. This thesis explores linguistic fuzzy sets and rule-based FLSs to provide understandable, transparent and simulatable assessments of driving risk. In addition, we explore hierarchical FLSs [225, 226, 259] to optimise the rules and ensure the system is decomposable i.e., the ability to explain different components of a system [258]. Chapter 5 provides the rational for using a hierarchical rule-based inference system for predicting and understanding the level of driving risk.

(g) Evaluation of performance

The development of a reliable online driving risk system is crucial for ensuring trust, acceptance, and successful adoption among stakeholders, end-users and decision makers [260, 261]. With the lack of labelled driving risk datasets, measuring the reliability of the system is difficult. Therefore, there is the opportunity of investigating ways to measure the reliability and effectiveness of systems with the absence of labelled data. For instance, this thesis (Chapter 5) proposes two novel metrics based on user studies to measure reliability by comparing the decisions produced by the AI system with assessments provided by domain experts, taking into consideration imprecision of human perception and AI systems.

2.5 Summary

HGV driving is at the forefront of trade and commerce in every nation, as both private and public sectors rely on it for the delivery of goods and services. Due to their importance, there is a growing interest by the transportation community to identify risky driving behaviours and assess their impact on road safety.

This chapter reviewed psychological theories on driver behaviour i.e. Theory of Planned Behaviour, Risk Homeostasis Theory and Multiple Resource, to understand and identify characteristics of a real-world driving environment. The theories explain the following characteristics of the driving environment: (1) driver behaviour is multifaceted consisting of the interaction of multiple secondary tasks or driving behaviours; (2) external factors constantly interact with driving behaviours to impact driving risk; and (3) there is always some level of driving risk. The characteristics are illustrated by the theoretical framework presented in Figure 2.2, and which forms the basis for this thesis.

Furthermore, we identified four main types of risky driving behaviours, including drivers' physical distraction, aggressive driving styles, responses to emotion and fatigue. The chapter later reviewed intelligent data-driven algorithms employed in the literature to characterise and predict the driving behaviours. The algorithms were grouped into four categories: (1) unsupervised learning methods; (2) conventional supervised learning methods; (3) deep learning methods; and (4) fuzzy systems. The review was summarised in Table 2.2. The main opportunities and gaps identified in the literature for improving the characterisation of driving behaviours are: capturing uncertainties in driving features; modelling imprecision in the interpretation of driving behaviours; capturing temporal dynamics in behaviour; and developing privacy-preserving strategies to protect drivers' identities and sensitive information. Most importantly, assessing the interaction of the different facets of driver behaviour that occur simultaneously during driving.

Lastly, the chapter reviewed the literature on driving risk consisting of post-hoc analysis and data-driven intelligent assessment of driving risk. We identified several opportunities to improve the automatic assessment of driving risk, including incorporating contextual information and modelling imprecision in information and driver characterisation AI systems. In addition, little has been done in the literature with regards to the online assessment of HGV driving risk.

This thesis proceeds to explore intelligent approaches to improve the characterisation of the risky driving behaviours using three data sources i.e., unlabelled telematics data, labelled driving posture footage and labelled facial footage. This is the first aim of this thesis as its achievement will ensure that inputs into the driving risk system are accurate and reliable. The next chapter presents novel AI approaches for detecting the different driving behaviours.

Chapter 3

Intelligent Driver Characterisation

3.1 Introduction

Chapter 2 reviewed three key psychological theories to understand the characteristics of driver behaviour in a real-world driving environment (Section 2.2.1, page 7). Those theories revealed that commercial driving consists of constant interactions between drivers' multifaceted behaviours, feelings and actions, as well as the influence of external factors. Therefore, to achieve a reliable and realistic online assessment of heavy goods vehicle (HGV) driving risk, driving behaviours need to be accurately classified from multi-modal driver data streams. After that, the impact of the detected driving behaviours on road safety are assessed. An accurate and reliable characterisation of driving behaviours can be achieved by exploring the artificial intelligence (AI) approaches reviewed in Section 2.3 (page 12) for driver behaviour analysis. However, due to the complex, dynamic and multifaceted nature of driver behaviour, stand-alone data sources are insufficient for examining the different driving behaviours in accordance with the literature reviewed i.e., driving styles, affective states, and driving postures. For example, telematics data alone could be analysed to describe and predict driving styles as reviewed in Section 2.3.1, but the data cannot tell us much about the driver's affective or distracted states.

This chapter aims at achieving the above objective and answering the research question: how can the intelligent characterisation of driving behaviours be improved using AI methods? To answer this question, AI methods are investigated to improve the detection of driving behaviours from multiple sources of data. Three data sources are investigated in this chapter to achieve its objective i.e., telematics incident data, driver posture images, and facial images as shown in Figure 3.1. Telematics data that capture the manner by which drivers operate vehicle controls has shown to be effective in characterising driving styles as revealed in Section 2.3.1. Driver-facing cameras capture the physical and mental states of drivers. A well-positioned camera can capture both drivers' postures and their faces. However, due to the lack of footage that captures both drivers' postures and their faces, two additional data sources are explored i.e., driving posture images captured by side-view driver-facing cameras, and facial images captured by facial-view cameras. The telematics data is gathered and provided by Microlise [262], our industrial partner. This dataset captures driving incidents, such as, harsh braking and over speeding incidents. The data are explored to identify the core driving styles of HGV drivers and develop an intelligent approach to classify the occurrence of driving incidents into one of the driving styles. The driving posture image dataset (The American University in Cairo Distracted Driver dataset [172, 263]) is available online and consist of 10 different driving postures. The dataset is utilised to develop a deep learning approach to improve the prediction of driving postures. Lastly, the facial image database (Remote Collaborative and Affective Interactions database [264]) is also publicly available. The database consists of facial images and their respective affect continuous labels. The Remote Collaborative and Affective Interactions (RECOLA) data are also explored to introduce an intelligent approach of protecting drivers' facial identities as facial images are difficult to anonymise.

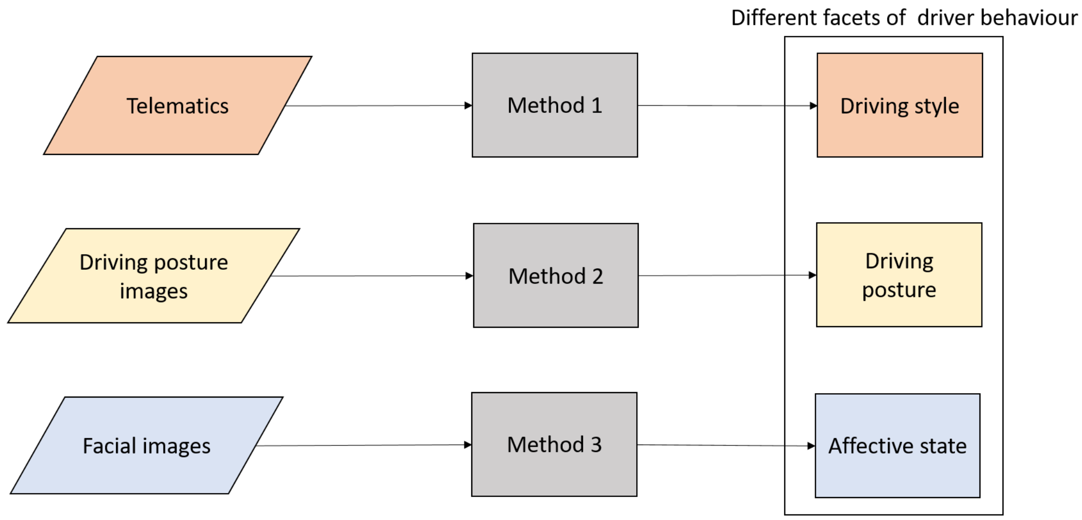


Figure 3.1: The diagram illustrates the data sources explored in this thesis to accurately detect the different facets of driver behaviour.

The chapter is organised as follows. Section 3.2 reviews the literature of characterising the different driving behaviours using AI approaches. Section 3.3 introduces our proposed methods for improving the characterisation of the different aspects of driver behaviour i.e., methods 1, 2 and 3 in Figure 3.1. The datasets used to evaluate the methods are described in Section 3.4 as well as the experimental design and evaluation protocols. Section 3.5 discusses the results of our experiments and opportunities for further improvement of HGV driver characterisation. Lastly, the chapter is summarised in Section 3.6.

3.2 Related Work

This section provides a review of the literature on the characterisation of driving styles, driver distraction postures and affective states using intelligent data-driven algorithms.

3.2.1 Driving styles

As mentioned in Section 2.2.2, drivers' aggressiveness is one of the main facets of driver behaviour that impacts driving risk [57, 58, 59, 60, 61]. Advances in wireless communication and sensors within vehicles have inspired the generation of data that captures the manner by which drivers operate vehicle controls, such as speed, acceleration, vehicle orientation and driving incidents. Researchers have explored different intelligent data-driven approaches to automatically detect and distinguish between driving styles, such as unsupervised learning [14, 57, 128, 129, 130, 131, 132], semi-supervised [129, 265, 266, 267] and supervised learning methods [45, 138, 139, 142, 143, 145, 146].

Constantinescu *et al.* [77] used an unsupervised learning method (Hierarchical Clustering Analysis) to identify six driving styles based on four telematics driving features i.e. speed, acceleration, braking and kinetic energy. The telematics data used in their study was collected from only 23 non-professional drivers. The authors employed Principal Component Analysis to interpret and distinguish between the driving styles using linguistic labels, such as 'very low', 'neutral', 'sudden' and 'high'. Even though their findings are important in understanding driving styles, their work is limited to the type and number of drivers investigated. Similarly, Figueredo *et al.* [241] employed an unsupervised learning method (ensemble of clustering techniques [268]) to uncover eight HGV driving styles in the United Kingdom (UK). The authors utilised a much larger and diverse dataset compared to Constantinescu *et al.* [77] (i.e., 1 year driving incident data from 21,193 HGV drivers). The authors developed an algorithm to interpret the driving styles using linguistic categories, such as 'low', 'moderate', 'high' and 'very high' number of incidents. The algorithm compares the median values for each variable of each driving style with the range of the entire data. However, similar to Constantinescu *et al.* [77] the driving styles identified are dependent on the data, which need to be large, diverse and comprehensive for reliable results. In addition, the driving styles are dependent on environmental conditions (contextual external information), which are not considered or included in the analysis.

With the scarcity of real-word labelled driver operation data and the difficulty to manually label large amounts of data, semi-supervised learning has been used to address these problems for driving style classification [129, 265, 266, 267]. The methods employ unsupervised approaches to group and label instances according to similar driving patterns. Subsequently, they employ supervised learning methods on the labelled data to predict the driving styles of unlabelled or new instances. For example, Wang *et al.* [266] proposed a semi-supervised method based on k-means clustering and support vector machines (SVMs) for driving style classification. Their study used a driving simulator for collecting driving data from 20 non-professional drivers driving on curvy roads i.e., speed, and throttle opening. The authors used a rule-based approach to provide subjective interpretations of driving styles i.e., normal and aggressive driving styles. Subsequently, Liu *et at.* [267] explored the opportunities of fuzzy sets in capturing data uncertainties injected by sensor readings. The authors employed fuzzy c-means to identify different driving styles while capturing uncertainties in driving data i.e., speed, acceleration. The data utilised in the study were collected from 51 non-professional drivers. The authors used the box-plots of features similar to Figueredo *et al.* [241] to interpret the driving styles as aggressive, general and conservative.

Recently, deep learning approaches have been applied on labelled driving data obtained under controlled experiments to classify driving styles [267, 269, 270, 271]. Controlled experiments or driving simulators can assist in labelling driving data as drivers are instructed to mimic different driving behaviours during operation. For instance, Bejani and Ghatee [269] trained convolutional neural networks (CNNs) using acceleration data collected from smartphones in controlled experiment to automatically recognise normal and aggressive driving styles. Similarly, Moukafih *et al.* [271] explored a publicly available driving data collected using smartphones under controlled experiments i.e., GPS, distance to following vehicle, speed and acceleration. The data were collected from six non-professional drivers performing three different driving behaviours (i.e., normal, drowsy and aggressive driving) on motorways and secondary roads. The authors developed a deep learning method based on long short term memory neural networks.

Furthermore, researchers have explored uncertainty models (e.g., fuzzy logic systems (FLSs) [217, 244, 272, 273, 274] and Bayesian networks (BNs) [156, 275]) to capture and model uncertainties in data injected by sensor readings, imprecision in the interpretations of driving styles and ambiguity in understanding the driving styles due to human subjectivity. For instance, Aljaafreh *et al.* [217] developed fuzzy sets to capture uncertainties in driving data i.e., acceleration and speed. The fuzzy sets coupled with predefined rules were used to develop a FLS to classify drivers into four driving styles i.e., below normal, normal, aggressive, and very aggressive driving styles. Subsequently, Cordero *et al.* [274] developed fuzzy sets to represent the manner of vehicle control (e.g., use of horn, control of steering wheel) and distraction states (e.g., gazing). They developed fuzzy rules to map the co-occurrence of the inputs to their respective driving styles i.e., ecological, normal and aggressive. Similar to Aljaafreh *et al.* [217] the fuzzy sets, rules and driving styles are fixed and pre-determined by the authors, and therefore, require further validation. In a different study, Han *et al.* [156] used Bayesian probability with kernel density estimation to extract discriminative driving features and estimate the probability of being aggressive or normal. They collected data from eight participants using a driving simulator. Features included speed, throttling, acceleration, position and steering angle.

The majority of the studies reviewed above for driving style characterisation utilise small telematics data with less than a hundred drivers. They employ single clustering or supervised learning methods, which could lead to unstable and model-specific driving styles. They also do not extract the relationships between the driving styles and driving features, which are important for improving risky driving styles and developing effective interventions. In addition, the reliability of assessing driving styles can be improved if uncertainties produced by imprecision in the description of driving features and driving styles are effectively represented and modelled.

To produce more reliable and interpretable HGV driving styles, this thesis introduces a hybrid fuzzy logic framework. The framework consists of an ensemble of clustering and supervised learning methods to uncover stable and reliable driving styles. Subsequently, the framework employs a data-driven FLS to model uncertainties in the description of driving incidents and driving styles. The FLS also extracts linguistic rules from the data, which represent the relationships between the occurrence of driving incidents and driving styles. The framework is applied on a large database of HGV telematics driving incidents to improve the characterisation of HGV driving styles. The database consists of four years of driving incidents from more than 2,000 HGV drivers, who completed more than 40 journeys yearly on different roads in the UK (a more detailed description of the database is provided in Section 3.4).

3.2.2 Driving postures

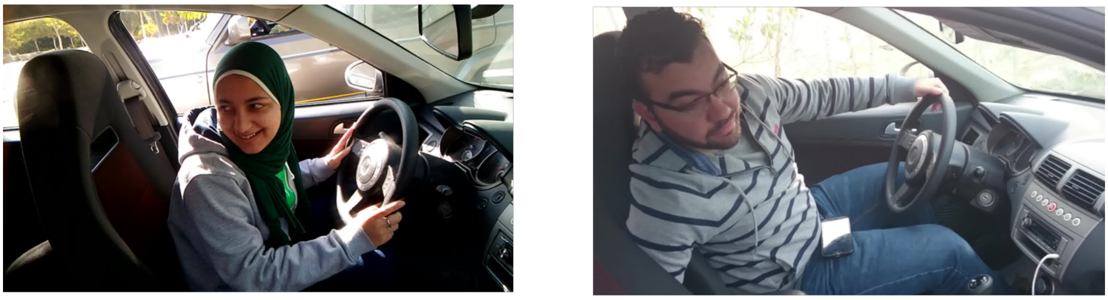
Another facet of driver behaviour identified in Section 2.2.2 that impact road safety is driving posture. Driving postures are categorised into 'safe' and 'distracted' postures [44]. Figure 3.2 shows images of 'safe' and 'distracted' postures extracted from the American University in Cairo (AUC) Distracted Driver Dataset. 'Safe' postures contain individuals in an alert driving position and looking ahead with hands on the steering wheel, while 'distracted' driving occurs when a driver engages in other activities that take their attention away from the road e.g., using the radio or in-vehicle technologies, cell phone usage, and looking at something outside the vehicle. These postures are difficult to detect using telematics data; however, they are easier to identify in images that capture the driver. Intelligent approaches that can automatically process images are therefore more suitable for detecting and distinguishing between driving postures. Those include computer vision approaches. This section reviews the literature on driving posture detection using driving footage to identify opportunities to improve driving posture characterisation.

Figure 3.2: Images extracted from the American University in Cairo Distracted Driver Dataset representing 'safe' and 'distracted' driving postures

Initial studies on driver distraction detection using intelligent methods and driving footage were based on handcrafted features and conventional supervised learning methods, such as Support Vector Machine (SVM) classifiers and ensemble of decision trees. For instance, Artan *et al.* [118] used a hand-crafted feature learning technique (known as deformable part model) to extract geometric features from driving posture images and employed SVMs to detect cell phone usage. The authors used a private database consisting of 1,500 images (378 drivers using cell phone and 1,122 drivers without using cell phone). Similarly, Berri *et al.* [46] explored SVMs for detecting cell phone usage in images of drivers. They used a hand-crafted feature learning technique to extract Haar-like-features for the identification of cell phone usage. Their method was evaluated on a database consisting of 100 images of drivers using cell phones and 100 images with no phone. Craye and Karray [45] explored hand-crafted features using AdaBoost and Hidden Markov models. Four feature sets were extracted from images of eight drivers: 1) arm position; 2) face orientation; 3) facial landmarks; and 4) gaze estimation and eye closure. The extracted features were concatenated into a seventeen-feature vector for each image frame. Subsequently, five postures were analysed: making a phone call, drinking, sending an SMS, looking at an object inside the vehicle and driving normally. The main challenge of hand-crafted feature learning techniques is that they require developers to be experts in image processing and the application domain.

With CNNs outperforming conventional supervised learning methods in image classification [43, 276, 277, 278, 279, 280], CNNs are now employed for driver distraction detection. For instance, Kim *et al.* [44] proposed a method of detecting driving postures using RestNet and MobileNet CNN models. Their study utilised a database of driver images consisting of 2,000 images in the training set and 2,000 images in the test set. The database includes two types of driving postures: looking in-front and not looking in-front postures. Their results on training models from scratch and using fine-tuned pre-trained models on distracted posture images of two drivers show that fine-tuned models significantly outperformed training from scratch for their database. Similarly, Yan *et al.* [43] examined CNNs to classify driving postures using pre-trained CNNs using three datasets: the *Southeast University Driving Posture* dataset [281], and two datasets developed by the authors called *Driving-Posture-atNight* and *Driving-Posture-inReal* datasets. Results show high classification accuracy with the three driving posture datasets which outperformed methods using hand-crafted features. Majdi *et al.* [42] combined CNNs with random forests (RFs) for the problem. Their model was trained on the AUC Distracted Driver dataset [172, 263], which consists of 30,000 images and 44 drivers. Their results show better performance compared to Support Vector Classifiers and CNNs without random forests. Eraqi *et al.* [172] proposed a weighted ensemble of CNNs using four different CNN architectures (i.e AlexNet , InceptionV3, ResNet and VGG-16 networks). They are trained on five different image sources of the AUC distracted driver dataset i.e. raw images, skin-segmented images, face images, hands images, and face and hands images. The results from the individual CNNs show the best accuracy when trained on raw images. The weights of the different CNNs are optimised using genetic algorithm and combined to improve performance. Similarly, Aljasim and Kashef [282] proposed an ensemble of CNN architectures that combined the outputs of two architectures to accurately distinguish between driving postures. The authors evaluated all combinations of ResNet, VGG16, MobileNet and Inception. Their results show that ResNet performed best for the individual architectures and ResNet-VGG16 increased the accuracy by 4%.

The majority of the CNN architectures explored in the literature are designed to capture spatial discriminative features. Therefore, driving postures with similar spatial representations are difficult for the CNNs to distinguish. For example, the driving postures in Figure 3.3 have similar spatial representations but represent different temporal dynamics. Intelligent approaches that capture differences between the spatial features (pixels) as well as temporal features could be explored to improve driving posture detection, such as recurrent neural networks (RNNs), attention-based CNNs and vision transformers.



'Talking to passenger' driving posture

'Reaching behind' driving posture

Figure 3.3: Sample images extracted from the American University in Cairo Distracted Driver Dataset with similar body movements but representing different driving postures

This chapter examines RNNs to model temporal dependencies between driving posture images and improve the accuracy of detecting driving postures. In addition, the memory cell of Long Short Term Memory networks (LSTMs) enables the networks to model long-term dependencies in sequential data as described in Figure 2.6 (page 20). LSTMs can also handle sequences of variable length by adjusting the number of time steps used in the computation, which is common in real-world driving as the duration of driving postures is not constant. As result of temporal and dynamic characteristics of driving, a novel hybrid deep learning architecture is introduced, which consists of CNNs to extract the spatial discriminative feature maps from a sequence of images and stacked RNNs to process the feature maps in a sequential manner and extract temporal discriminative features. Further details are provided in Section 3.3.2. Due to data privacy challenges in obtaining driving posture images of HGV drivers, this thesis utilises the large and diverse publicly available AUC distracted driver database. This thesis also addresses the issue of individual data protection in Section 3.3.3.

3.2.3 Driver affective states

In Section 2.2.2, driver emotion and fatigue were identified as major affective states that impact driving performance. These states are more identifiable in facial images than in telematics data or driving posture images. Similar to driving posture detection, the literature has shown that CNNs are most suitable to detect affect in facial images due to their ability to extract spatial discriminative features that distinguish the states. However, as mentioned in Section 2.3.3, analysing facial images result in several data privacy and driver protection concerns. This section reviews the literature on affect recognition using facial images and deep learning to identify opportunities to improve the detection of emotion and fatigue while safeguarding driver's privacy.

Deep learning (e.g., CNNs and RNNs) has been widely explored to predict human affective states [193, 194, 195, 198]. For instance, Tzirakis *et al.* [193] trained a convolutional recurrent neural network where the spatial features of a sequence of images are extracted using CNNs with fully connected layers, subsequently processed using LSTMs. The authors utilised a publicly available affective dataset called Remote COLlaborative and Affective interactions (RECOLA) [264]. It contains continuous response variables for valence and arousal affective states. Valence represents the intensity of a human's emotion between sad and happy, where -1 represents sad and +1 represents happy. Arousal is the intensity between calm and excited, where -1 represents calm and +1 represents excited. Their results reported lower performance in predicting arousal compared to valence using facial images. Lee *et al.* [194] developed an ensemble deep learning architecture that combines features extracted using 3D CNNs with the outputs of convolutional LSTMs to predict valence. They also utilised RECOLA but reported lower prediction performance of valence compared to Tzirakis *et al.* [193]. Subsequently, Lee *at al.* [283] combined the raw facial images with depth and thermal representations of the images to improve facial affect recognition using a deep learning architecture consisting of CNNs and attention-based LSTMs. Their results on the RECOLA database showed an improvement in performance with multi-modal data fusion compared to single modalities. Akhand *et al.* [284] recently explored transfer learning to improve prediction accuracy. They developed a deep CNN and replaced some layers of their network with layers from a model trained on the ImageNet. They evaluated the performance of their approach using different models trained on the ImageNet (e.g., VGG, ResNet, Inception and DenseNet). However, their approach was evaluated on databases with discrete response variables and therefore, cannot be compared to the other studies that used continuous response variables. This review of affect recognition studies that utilise deep learning models reveals the best performance (according to concordance correlation coefficient (CCC) metric) on the RECOLA database (same database explored in this thesis) as 0.620 for valence and 0.464 for arousal. These results show the need for more work to be done to improve prediction accuracy to a performance close to 1.0, which represents perfect agreement between actual and predicted values. In addition, the review showed that more studies focused on valence prediction. However, arousal is as important as valence for predicting fatigue as fatigue is highly correlated to arousal [285, 286].

To protect users' facial identity and still predict affective states, researchers have explored anonymised facial Action Units (AUs) extracted from facial expressions in images [196, 197, 198, 199]. These anonymised facial features (AUs) represent human-observable facial muscle movements that estimate the intensity of facial movements using facial landmarks. For example, AUs 12 (raising lip corners), 15 (lowering lip corners) and 20 (lip stretch) can be estimated using the facial landmarks on the lips. Figure 3.4 shows 98 points or landmarks on the face developed by Wu *et al.* [3] that represent important features of the human face and could be utilised as anonymised facial features. Han *et al.* [199] proposed a ensemble approach that combines the outputs of support vector regressor (SVR) and LSTMs to predict valence and arousal affective dimensions. The authors explored facial landmarks extracted from the RECOLA image database to develop a privacy-preserving

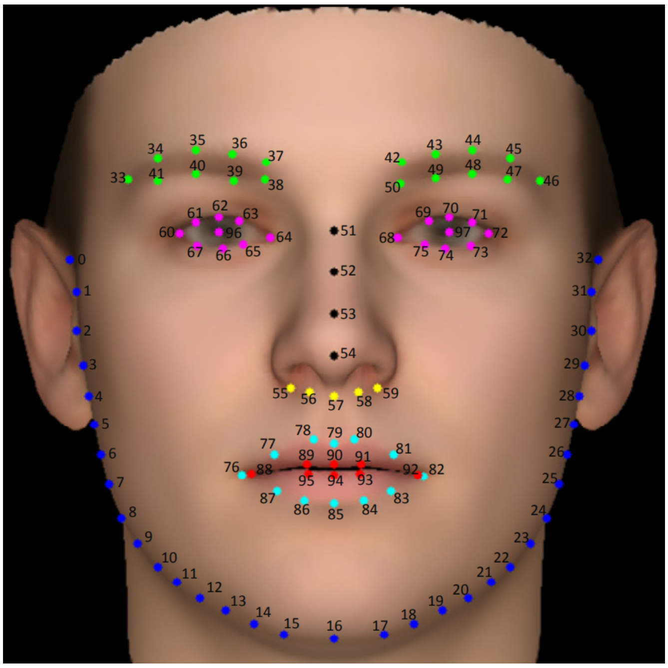


Figure 3.4: 98 facial landmarks aligned on the face of an avatar depicted from Wu *et al* [3]

approach. Their experiments showed best CCC performance of 0.394 for valence and 0.265 for arousal. Another example of a study that explored anonymised facial feature is Ortega *et al.* [196]. The authors developed a SVR to process facial landmarks extracted from the RECOLA image database. Their experiments show best CCC performance of 0.344 for valence and 0.337 for arousal. Similar to facial affect recognition, the performance of current approaches for processing facial features require significant improvement. In addition, even though it is difficult to reconstruct faces from the AUs or facial landmarks, auto-encoders and generative adversarial networks that can reconstruct images from latent features could make face reconstruction possible [287, 288, 289, 290]. As a result, there is still need to explore intelligent approaches that prevent complete access to users' identities and sensitive information.

This chapter explores a federated learning (FL) implementation of the hybrid deep learning architecture introduced for driving posture detection to detect valence and arousal affective states while safeguarding the identity and privacy of users. In FL, models are trained on users' data within their local machines and transferred to a central machine for aggregating [200]. Further information about our approach is described in Section 3.3.3.

3.3 Frameworks for Driver Behaviour Detection

This section introduces our proposed methods for improving the characterisation of driver traits i.e., driving styles, driving posture and affective states, as shown in Figure 3.1. First, a novel hybrid fuzzy framework is presented that consist of clustering coupled with supervised learning and FLS. The framework captures uncertainties in telematics data, identifies robust and core HGV driving styles, and describes the driving styles in terms of driving incidents using understandable linguistic terms and fuzzy rules. Subsequently, a novel hybrid deep learning framework for distraction posture detection is introduced, which consist of CNNs coupled with LSTMs. The CNN architecture extracts discriminative spatial feature maps from images, which are later sent to the LSTM architecture. The LSTM architecture captures the sequential dependencies among the spatial features. Lastly, a federated learning implementation of the hybrid deep learning framework is proposed for predicting valence and arousal affective states while protecting users' privacy and identity.

3.3.1 A hybrid fuzzy framework for driving style characterisation

The framework is shown in Figure 3.5 and each stage is described in detail below.

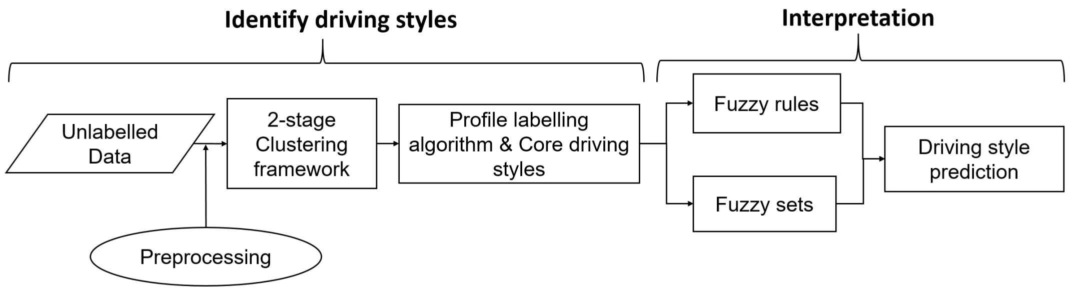


Figure 3.5: A hybrid fuzzy framework that consists of clustering coupled with supervised learning and fuzzy logic inference for driving style characterisation.

Stage 1: Data Pre-processing

Data are preprocessed to manage missing values, outliers, irrelevant features, and highly correlated features. Next, they are normalised to ensure unbiased results e.g. the number of driving incidents are divided by the total driving time to capture the effects of journey length as drivers who complete longer journeys are prone to more incidents. Instance selection can also be performed for removing noisy instances especially for obtaining reliable data patterns. For example, only drivers who are present across all years could be considered for obtaining more reliable driving styles.

Stage 2: A 2-stage clustering framework

To uncover stable and reliable driving patterns, a 2-stage clustering framework introduced by Agrawal *et al.* [268] is employed. The framework uses a consensus of clustering methods and an ensemble of classification techniques to produce robust and stable data labels. In the first stage, multiple clustering algorithms are run to group the data. Using a consensus approach, most of the data are assigned to one of the identified labels, while some of the data remain unclustered due to lack of consensus among the clustering algorithms. In the second stage, multiple classification algorithms are trained on the clustered data from stage 1 of the clustering framework. The trained models are applied on the unclustered data to assign them to one of the labels using a ensemble approach. This framework provides an automatic and systematic approach to assign robust labels to instances of datasets especially in domains where data patterns change frequently e.g. in commercial driving where driving styles are frequently affected by changes in external conditions and customer demand. The framework showed state-of-the-art performance in numerous datasets e.g. Iris, Wine, Ecoli, Wisconsin Breast Cancer, and Dermatology [268].

Stage 3: Profile labelling algorithm and core driving styles

In order to understand the driving patterns obtained by the 2-stage clustering framework in accordance with driving features, a well-defined profile labelling algorithm introduced by Figueredo *et al.* [241] is employed. The algorithm provides a systematic approach to describe data patterns or clusters. This is done by labelling each feature in each cluster using three linguistic terms; 'low', 'moderate' and 'high', to describe their occurrence. The algorithm calculates the median for each feature of each cluster, and compares the calculated median with the five-number summary values of the feature for the entire dataset as shown in Algorithm 1. The labels assigned to the features in clusters are examined by domain experts to provide human-understandable descriptions of the clusters. That is, instead of the incomprehensible cluster labels e.g., 1, 2, 3 etc, domain experts can use the feature labels in the different clusters to provide meaningful and understandable labels e.g., calm, reckless and aggressive driving styles.

Algorithm 1: Label Clusters

inputs : Box-plot of $Variable_i$ in $Cluster_j$, box-plot of $Variable_i$ in entire dataset.

output: $label_i$ for $Variable_i$ in $Cluster_j$

```
foreach Cluster in Cluster_j do
    foreach Variable in Variable_i do
        if Median(Variable_i inCluster_j) ≤ Median(Variable_i inAllData)
         then
            label_i ← "Low";
        else if Median(Variable_i inCluster_j)
         IsBetween(Median(Variable_i inAllData)
        and ThirdQuartile((Variable_i inAllData))) then
            label_i ← "Moderate";
        else if Median(Variable_i inCluster_j) IsBetween
         ThirdQuartile(Variable_i inAllData) and
         Maximum(Variable_i inAllData) then
            label_i ← "High";
```

To produce core driving styles from the driving clusters, similar clusters are merged. That is, clusters which differ by *only one* label (i.e. low, moderate or high) in *only one* of the driving incidents are merged, while those that do not follow this rule form standalone driving styles. By defining core driving styles, a more general and robust characterisation of driver behaviour can be achieved.

Stage 4: Generating fuzzy sets and rules

In this stage, feature (input) membership functions (MFs) consisting of the fuzzy sets (FSs) 'low', 'moderate' or 'high' are generated from the labelled data obtained from the previous stages using the five-number summary of the features similar to the profile labelling algorithm. The five-number summary consists of minimum (min), first quartile (Q_1), median (Q_2), third quartile (Q_3) and maximum (max) values. Z and S-MFs are utilised for 'low' and 'high' FSs as they are most suitable for representing the extreme boundaries with maximum degrees or likelihood of membership. The Z-MF is generated using the minimum and median values and the S-MF employs the median and maximum values. For 'moderate' FS, Triangular MFs are employed as they are efficient in representing intermediate FSs. First quartile, median and third quartile values are utilised to construct triangular-MFs. Figure 3.6 shows an example of membership functions generated from the five-number summary statistics of a feature i.e. min=0, Q_1=25, Q_2=50, Q_3=75, and max=100.

FSs for the driving styles are also generated, such as 'calm', 'moderate' and 'aggressive' driving styles. The FSs are equally distributed between 0 and 100 to ensure a smooth transition between driving styles, where 0 represents calm driving and 100 represent aggressive driving.

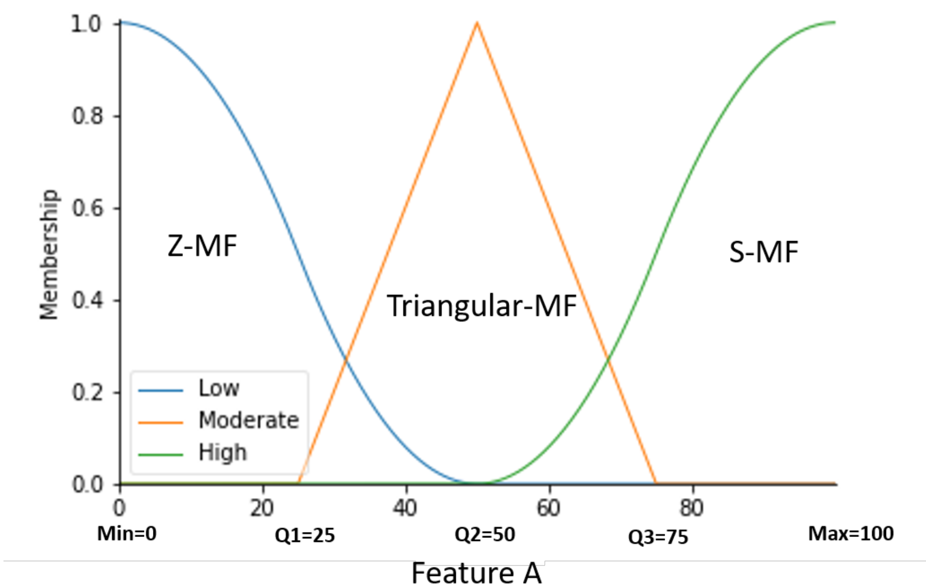


Figure 3.6: A diagram to illustrate the generation of membership functions using the five-number summary of a feature.

A rule generation method (e.g. Wang-Mendel rule generation method [291]) is applied on the labelled data to automatically generate IF-THEN rules that map the driving patterns to the driving styles. This method generates rules for all possible combinations of input FSs and enables the system to adapt to changes in data patterns, especially for dynamically changing environments e.g., HGV driving environment.

In a labelled dataset, the inputs and output (*profile*) pairs are denoted as:

$$\begin{array}{l} (x_1^1, x_2^1, x_3^1, \ ... \ x_n^1; y^1), \\ (x_1^2, x_2^2, x_3^2, \ ... \ x_n^2; y^2), \\ \quad \vdots \qquad\qquad\qquad\qquad \vdots \\ (x_1^m, x_2^m, x_3^m, \ ... \ x_n^m; y^m), \end{array} \tag{3.1}$$

where m is the number of input-output data pairs and n is the number of features.

Each input-output pair is assigned to the corresponding input and output MFs to produce 'IF-THEN' rules and the number of occurrences of each rule (weight). The rule weights are utilised to implement a rule reduction procedure for conflicting rules. The rules could be further improved by domain experts due to their simple linguistic representation.

The generation of input FSs and rules from data enables the system to be flexible to changes in driving patterns. For example, if a new road is constructed that causes all drivers to generate more incidents than usual, this will be captured by the data-driven MFs as a general increase in incidents. Therefore, there will be no shift of FSs. Whereas, if some drivers suddenly become more aggressive, the MFs will capture this change in behaviour and the FSs will shift accordingly and the rules adapted to capture the relationship between the new driving pattern and the aggressive driving style.

Section 3.4.2 elucidates the different membership functions and rules generated from the data using the proposed data-driven fuzzy approach.

Stage 5: Driving style prediction

Mamdani rule-based fuzzy inference method is used to predict the core driving styles of input driving patterns i.e., driving incidents obtained from sensors. Mamdani inference converts crisp input values into FSs (fuzzification) using the input MFs; computes rule strengths using 'AND' (minimum of membership degrees), 'OR' (maximum of membership degrees) and 'NOT' (1 minus membership degree) [292] operators; and aggregates the rule strengths using aggregation operators. The aggregated rule strengths represent the likelihood of driving styles and the resulting IF-THEN rules represent meaningful mappings of driving patterns to driving styles.

3.3.2 A hybrid deep learning framework for driving posture detection

The proposed hybrid deep learning framework for driving posture detection consists of CNNs coupled with LSTMs. The CNN architecture consists of deep convolutional layers to extract discriminative spatial feature maps from images, which are later sent to the LSTM architecture. The LSTM architecture consists of bi-directional recurrent neural networks to capture the sequential dependencies among the spatial features and fully connected neural networks to predict the driving posture. The framework is shown in Figure 3.7 and each stage is described in detail below.

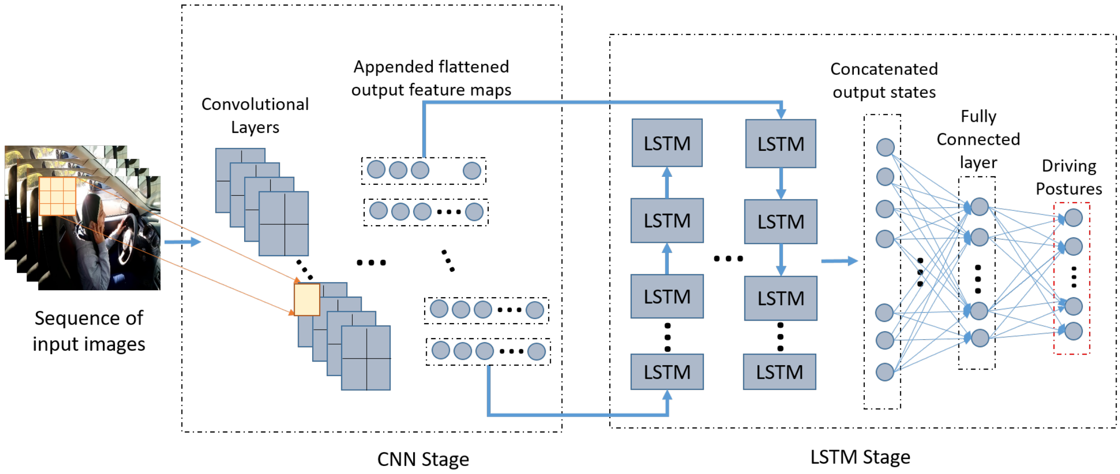


Figure 3.7: A hybrid deep learning framework to predict driving postures in images

Stage 1: Convolutional layers

This stage consists of convolutional layers to extract spatial features from a sequence of images representing different driving postures. First, images are transformed according to the requirements of the CNN architecture implemented and the pixel values are normalised to speed up model optimisation. For large and diverse datasets with several thousands of images for each driving posture, multiple convolutional layers are stacked to extract feature maps from the normalised data. The different components of a typical convolutional layer are described in Section 2.3.3 (page 17). In the case where there are only few images with very similar driving postures, existing deep CNN architectures that have been trained on image classification tasks (i.e., pre-trained CNN architectures) are utilised to take advantage of their optimised model weights. Some popular state-of-the-art CNN architectures are employed to evaluate the framework as described in Section 3.4.

The feature maps produced by the last convolutional layer in the architectures represent important spatial information for distinguishing between the driving postures e.g., head, body and hand positions. In order to capture the temporal differences between different driving postures, the feature maps are are stacked up and flattened. The flattened feature maps for each image in a sequence are stacked up (appended) to form a sequence of spatial features and fed to the LSTM architecture in the next stage.

Stage 2: Long short-term memory networks

The LSTM architecture consists of stacked recurrent neural networks that process the feature maps in a sequential manner to extract the sequential dependencies among the images that differentiates driving postures. This is done in both the forward (i.e. first to last feature map in the sequence) and backward directions (i.e. last to first feature map in the sequence) to increase the amount of information captured about the sequence. The stacked recurrent networks send the learned information from previous steps in the sequence to subsequent steps through multiple hidden layers. The output (temporal features) from the last recurrent cell in the forward direction is merged with that from the backward direction, and passed to a fully connected layer to classify the images. The recurrent networks share spatial information from previous feature maps using the cell (C_t) and hidden states (h_t), as follows:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c X_t + U_c h_{t-1} + b_c) \tag{3.2}$$

$$h_t = o_t \odot \tanh(C_t) \tag{3.3}$$

Where C_{t-1} is the spatial information of the previous flattened feature maps in the sequence, f_t is the forget gate's activation function of the current feature map, i_t is the input gate's activation function of the current feature map, o_t is the output of the current feature map representing both spatial and temporal information from previous feature maps, and $tanh$ is an activation function. W, U, and b represent the weight matrices and bias vectors which need to be learned during training.

3.3.3 A federated deep learning framework for affect recognition

The hybrid deep learning framework introduced in Section 3.3.2 is implemented in a federated learning fashion to predict human attributions of valence and arousal using facial images. Facial images consist of facial expressions that could be used to predict drivers' emotions and level of fatigue. The federated learning approach processes users' facial images at their local machines and sends their trained models to the central processing module for aggregation, as shown in Fig. 3.8. The local and central processing machines have the same model implementation to enable easy aggregation of the trained models. Different aggregation or information fusion techniques [293] exist to combine the locally trained models, such as mean, weighted average, minimum, maximum amongst others. The deep learning framework implemented at the local and central processing machines consists of CNNs coupled with recurrent networks to learn the spatio-temporal dynamics in sequences of images and for combining the model weights, respectively.

The training process occurs across multiple local machines, each containing an individual's images. After each training iteration or pre-defined time step, the central machine receives trained model weights from the different local machines. Users' images do not leave their local machines, thereby, preventing any access to users' data. In the central machine, the weights are aggregated and the aggregated weight is sent back to the local machines to update their weights for the next training iteration. The processes are repeated throughout the training stage i.e., local training, central aggregation and local weight update.

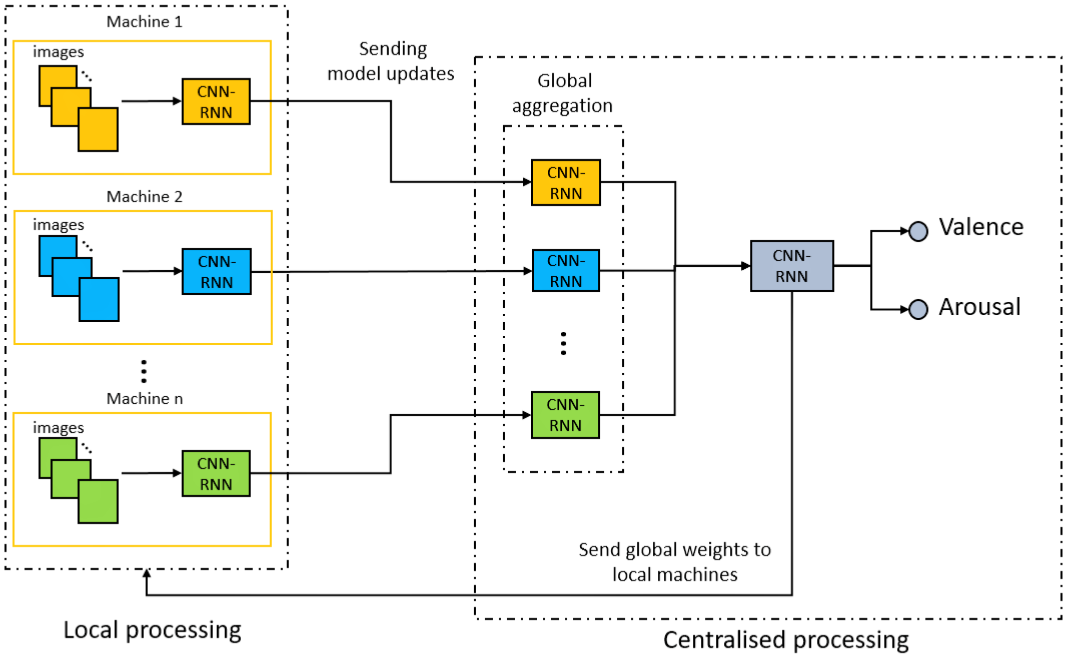


Figure 3.8: A diagram representing the different steps of our proposed federated deep learning approach for affect recognition.

3.4 Experiments

This section first describes the three datasets (i.e., HGV telematics incident data, AUC driving posture database, and RECOLA) acquired for evaluating the methods presented in the previous section (Section 3.3). Subsequently, the experimental configurations for the different methods and evaluation protocols are presented.

3.4.1 Datasets

The three datasets are described below:

Database 1: Telematics driving incident datasets

HGV telematics driving incident data is collected from Microlise [262], our industrial partner. Microlise employs telematics solutions for driving and vehicle data collection; and they use this information to define strategies for incident prevention and to promote better practices in the HGV industry. Information produced by their telematics solutions are transmitted and collected from the HGVs using a Controller Area Network (CAN) bus. The data consist of aggregated driving incidents for each driver who completed a minimum of 10 journeys per quarter each year. The data was collected between the first of January and the thirty first of December for the years 2014, 2015, 2016, and 2017. The data was split by clustering drivers' average daily mileage into three datasets i.e., short, medium and long mileage datasets. Short mileage drivers completed journeys with a daily average mileage up to *136.70* miles, medium mileage drivers had an average daily mileage between *136.70* and *217.48* miles, and long mileage drivers covered an average daily mileage more than *217.48* miles. In total, 2,253 HGV drivers for short, 3,776 for medium and 3,616 for long average daily mileage subgroups are considered across the four years. The datasets consist of four driving incidents: frequency of Harsh Braking (HB) events, Over-Speed (OS) duration in seconds, Excessive Throttle (ET) duration in seconds and frequency of Over Revving (ORev) events of HGV journeys completed in the United Kingdom. Those incidents were chosen because they are the most relevant driving incidents present in all of the HGVs, which are related to road safety. It is important to mention that these datasets are currently the largest telematics datasets (more than 2000 drivers) reported in the HGV industry with driving incidents collected for the longest period (4 years).

Database 2: The American university in Cairo distracted driver datasets

The AUC distracted driver database [172, 263] is the largest, most comprehensive publicly available database for driving posture characterisation. The database captures most real-world driving postures (up to 10 postures): safe driving (c0), text right (c1), right phone usage (c2), text left (c3), left phone usage (c4), adjusting radio (c5), drinking (c6), reaching behind (c7), hair or makeup (c8), and talking to passenger (c9). Figure 3.2 shows sample images of the different types of distracted postures found in the database. The driving posture images in the database were captured using an ASUS ZenPhone rear camera (Model Z00UD), with image sizes of 1080 * 1920 pixel. The images were captured from 44 participants in experimental settings. The database was later split into training and test sets. The training set consists of images from 38 participants and images of the remaining 6 participants formed the test set. Table 3.1 shows the number of images in the training and test sets for each driving posture. The database is not well balanced with 'safe driving' having the highest proportion of images.

Table 3.1: Number of images in each driving posture for training and test sets of AUC distracted driver Database

Types of driving postures	Number of images in training set	Number of images in test set
c0	2,440	266
c1	1,305	133
c2	862	114
c3	744	100
c4	950	90
c5	753	90
c6	733	63
c7	691	63
c8	698	66
c9	1,379	138

Table 3.2: Images of different types of distraction postures extracted from the American University in Cairo Distracted Driver Dataset.

Database 3: Remote collaborative and affective interactions datasets

The RECOLA database [264] is a popular and comprehensive affective database with continuous response variables (i.e. valence and arousal). The database consists of videos, AUs, audio, ECG and EDA datasets for 23 participants. The data was collected during spontaneous and naturalistic interactions between the participants when performing collaborative tasks. The database also contains the ground truth continuous labels for valence and arousal that range from -1 to +1. The annotations were carried out by six annotators. The annotators looked at each frame and provided their subjective opinion (in the continuous scale from -1 to +1) about the valence and arousal affective states of the facial expressions in the frames. They were properly trained to understand the task and annotation tool. Figure 3.9 shows the valence and arousal dimensions and some sample affective states that can be extracted from the dimensions [276]. This thesis utilises the facial images extracted from the RECOLA video dataset with a frame rate of 25fps and AUs extracted from the facial images. A total of 7,500 images per participant was extracted, and 15 AUs extracted from each image. In addition, movements of the face in X-Y-Z directions (i.e. pitch, roll and yaw respectively), the mean and standard-deviation of the optical flow in the region of the face, and changes of the AUs, facial movements, mean and standard deviation of the optical flow from the previous timestamp (delta coefficients) were computed and added to the 15 AUs to produce a total of 40 facial features per image. Figure 3.3 shows sample images with different valence and arousal labels extracted from the database.

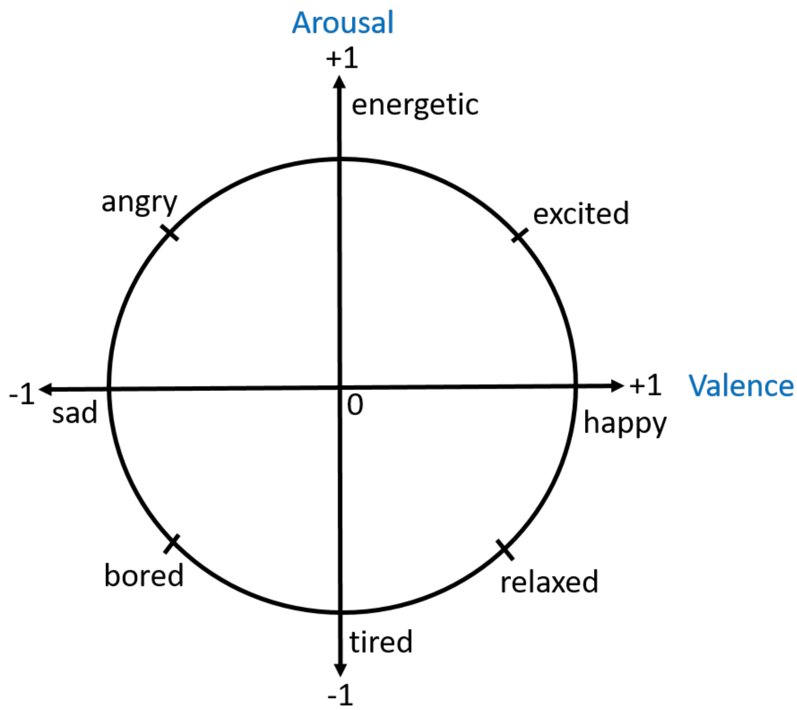


Figure 3.9: Two-dimensional description of affect using continuous scales

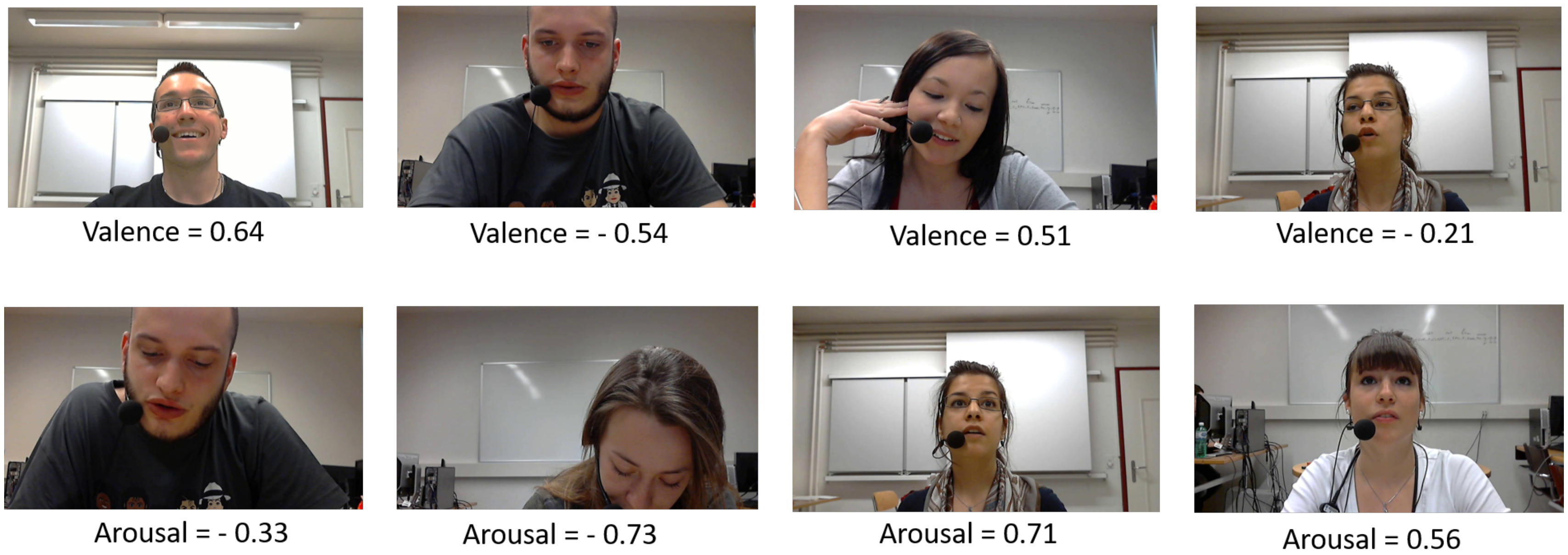


Table 3.3: Sample images with their labels of valence or arousal extracted from the RECOLA database.

3.4.2 Experimental settings

This section presents the models implemented in the various stages of the frameworks introduced in Section 3.3 and their configurations. The section also presents state-of-the-art models explored for comparison with the frameworks.

Method 1: Hybrid fuzzy framework

The consensus clustering stage consist of K-Means (KM) [294] and Partitioning Around Medoids (PAM) [295]. Calinski and Harabasz [121], and Hartigan [122] cluster validity methods are applied over a range of number of clusters varying from 2 to 20 to determine the ideal number of clusters in the data. For the second stage of the 2-stage clustering framework, Support Vector Machine (SVM), Random Forest (RF) and Multi Layer Perceptron (MLP) are employed as an ensemble on the clustered data. Due to the large size of the telematics data and hyper-parameter set, the models are fine-tuned using random search hyper-parameter optimisation [296]. The optimised hyper-parameters for the classifiers are shown in Table 3.4.

For the fuzzy logic stage of the framework, the five-number summary statistics for each driving incident are used to construct their respective membership functions (antecedents MFs). Figure 3.10 shows the membership functions for the four driving features of short daily average mileage subgroup generated from their summary statistics as described in Section 3.3.1. The consequents of the FLS are the core driving styles obtained from the 2-stage clustering framework. The consequent membership functions are equally distributed between 0 and 100, which represents the level of aggressive driving.

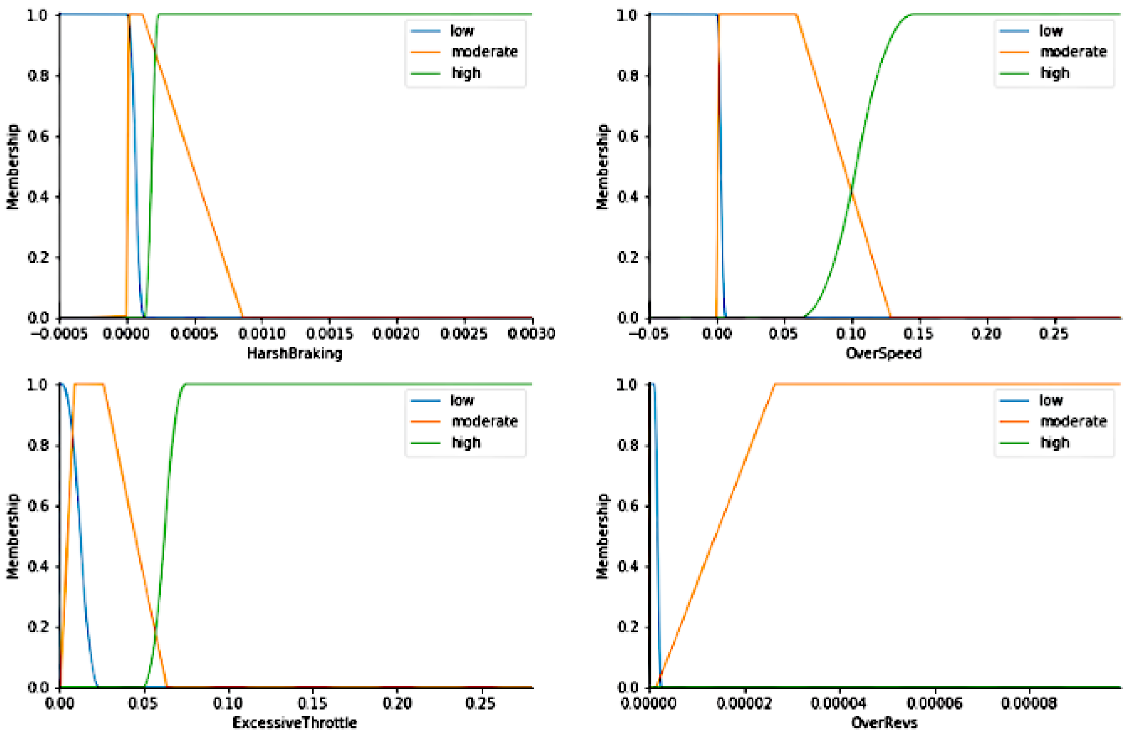


Figure 3.10: Driving incidents membership functions for short mileage drivers

Subsequently, the Wang-Mendel [291] is employed to generate the IF-THEN fuzzy rules. After generating the MFs and consequents from the data, each input-output pair in the data is fired on the MFs to produce IF-THEN rules consisting of input and output FSs. The occurrence of each rule from the data corresponds to the rule's weight. The rule weights are utilised to implement rule-reduction procedure on conflicting rules. Table 3.9 shows the rules generated from short mileage incident dataset using the Wang-Mendel approach and the MFs presented in Figure 3.10.

Lastly, Mamdani rule-based inference method [215] is employed for predicting the likelihood of driving styles by using the rules generated by the Wang-Mendel approach and fuzzy MFs generated from the five-number summary statistics. The inference process is described in Section 2.3.4 (page 21).

To compare the classification performance of our framework with other classifiers, the Mean Of Maxima (MOM) defuzzification [297] is utilised to reduce the output fuzzy sets to a single class. MOM is employed because it selects the class corresponding to the maximum value of the membership function i.e., the class with the most likelihood of occurring. The classification performance of our framework is compared with the individual classifiers used in the ensemble (i.e., RF, SVM, MLP), trained on the labelled datasets obtained after employing the profile labelling algorithm described in Section 3.3.1.The labelled incident datasets (i.e., short, medium and long mileage datasets) is split into training and test sets using a 70:30 ratio.

Table 3.4: Hyper-parameters of classifiers employed in stage 2 of the 2-stage clustering framework on short, medium and long mileage telematics datasets.

ML Approaches	Hyper-parameters	Short	Medium	Long
Random Forest Classifier	Number of Trees	200	300	300
	Max depth	20	40	40
	minimum samples for node	1	2	2
	Minimum samples to split	2	5	10
Support Vector Machine	Kernel	Radial basis	Radial basis	Radial basis
	Regularisation	10	1	1
	Kernel coefficient	1	1	1
MultiLayer Perceptron	Hidden Layer Size	150	100	200
	Activation function	ReLU	ReLU	ReLU
	Optimiser	Adam	Adam	Adam
	L2 Regulariser	0.001	0.001	0.01
	Learning Rate	0.001	0.001	0.001

Method 2: Hybrid deep learning framework

Inception-V3 CNN model [177] is pretrained on the ImageNet database and applied in the CNN stage of our framework due to its better performance (f1-score versus training time) compared to other CNN models on the AUC database as shown in Figure 3.11. Inception showed comparable performance with DenseNet with far lower training time. For the LSTM stage, LSTMs [182], Gated Recurrent Units (GRUs) [298], and bi-directional versions of the networks are implemented and evaluated.

To demonstrate the good performance of our framework in predicting driving postures, the performance of our framework is compared with state-of-the-art CNN

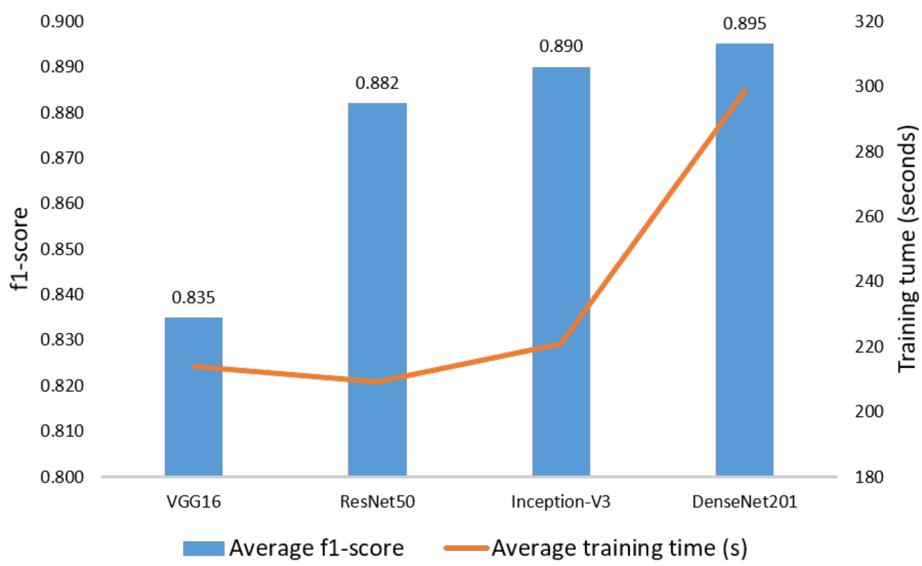


Figure 3.11: Average f1-score versus training time of popular pre-trained ImageNet convolutional networks on the AUC distracted driving posture datasets

models (i.e., VGG [176], Inception [177], ResNet [178] and DenseNet [179]), and past studies that explored deep learning methods on the AUC database. The CNN models have the following features:

- **VGG:** VGG was introduced in 2014. The network is very deep with thirteen convolutional layers and three fully connected layers.
- **Inception:** Inception was introduced in 2014 with its main motivation to deal with the uncertainty in choosing the kernel size of convolutional layers. The network uses multiple kernel sizes in each convolutional layer to complement the advantages of different kernel sizes while avoiding deeper architectures which can lead to overfitting.
- **ResNet:** ResNet was introduced in 2015 to handle the vanishing gradient problem [180]. The network solves this problem by introducing skip connections between layers. This allows the gradients to flow properly through the skip connections to any other earlier layer.
- **DenseNet:** DenseNet was introduced in 2017 as an upgrade of ResNet. In DenseNet, each layer connects to every other layer i.e., the input of any layer is the concatenated feature maps of all subsequent layers. This reduces the number of parameters compared to ResNet, however, very deep DenseNet variations have been recently developed that have very large number of parameters.

The AUC training set (data from 38 drivers) is split into a new training set with 35 drivers' data and a validation set with 3 drivers' data. The validation set is used to fine-tune the hyper-parameters of the models. The learning process is stopped when the validation loss stops decreasing for 5 consecutive epochs. Simple networks are first implemented and their performance improved by evaluating more complex configurations. The different hyper-parameter sets evaluated to improve performance are described below.

The models are evaluated using popular learning rates between 0.01 and 0.0001. Performance with 'no', 0.5 and 0.2 dropouts are compared as these dropout values have proven to significantly improve performance [299]. Dropouts were not implemented in the CNN models because already existing CNN networks were utilised. By considering the number of images in the datasets (less than 15k images) and the depth of the networks (very deep networks), small batch sizes of 8, 16, 32 and 64 were evaluated. To prevent overfitting in the recurrent networks, small number of stacked recurrent layers between 1 and 4, and hidden units in the range 16 to 128 with increments of 16 were evaluated. The input features of the recurrent networks are chosen between 64 and 256 with increments of 64, due to the size of the flattened output feature maps of the CNNs. The sequence length of the recurrent networks varied according to the number of the images for each driving posture sequence. Table 3.5 shows the best hyper-parameter values after evaluating the set of hyper-parameter values.

Lastly, the AUC test set consisting of 6 drivers was used to evaluate the performance of the optimised models. Each model is trained 5 times and the Average Accuracy (AA), Average F1-score (AF) and the Average Inference Time (AIT) computed on the test set. The experiments were executed on a graphics processing unit (GPU) using 4 CPU cores and 6GB RAM. Our code was implemented in Pytorch with an epoch size of 50 for each experiment.

Method 3: Federated deep learning framework

Similar to the hybrid deep learning framework employed for detecting driving postures, bi-directional GRUs and LSTMs are explored in the recurrent stage for detecting affect. However, a shallow ResNet network (i.e. ResNet18) is employed in the CNN stage of the framework as the best performance on the RECOLA was achieved using ResNet network [193].The ResNet-BiLSTM/GRU model is implemented on the local machines as well as the central processing machine. In a real-world driving setting, the local machines are processors found in the vehicles where the images are captured and locally stored, while the central processing machine is found on the deployment environment e.g. on-premise, and cloud servers. Each local machine or device trains a local version of the model on its own data and sends their trained models (weights and biases) to a central server for aggregation. The central server aggregates the models to update the global model and sends the updated global model back to the local machines, which use it to update their own local models and continue training. A federated deep learning approach is a method of training machine learning models using data that is distributed across multiple devices or locations, without the need to centralise the data in one location. This approach is often used when data privacy or security concerns prevent centralised training or when the data is too large to be centralised. The main advantage of the federated deep learning framework is that the central processing machine trains a global model without access to the data as data are stored and processed in the local devices. In addition, it can enable the training of models on data that is distributed across

Table 3.5: Hyper-parameter configuration of convolutional and recurrent networks on AUC distracted driving posture training dataset

	VGG-19	**Densenet-201**	**Resnet50**	**InceptionV3**	**InceptionV3-LSTM**	**InceptionV3-GRU**	**InceptionV3-BiGRU**	**InceptionV3-BiLSTM**
Batch size	32	32	32	32	16	16	8	8
Learning rate	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Input features	299*299	224*224	224*224	299*299	256	256	64	64
Hidden layer units	/	/	/	/	64	64	128	128
Number of layers	/	/	/	/	1	1	1	1
Dropout	/	/	/	/	No	No	No	No

multiple machines when data are too large to be centralised.

To validate the federated framework, its performance is compared with another privacy-preserving strategy i.e., processing anonymised facial features (AUs), and the conventional facial recognition strategy i.e., centralised processing of facial images. In addition, the performance of the framework is compared with other studies that employed machine learning methods on the RECOLA image and AU datasets for affect recognition.

The valence and arousal ratings provided by the six annotators are averaged to produce the ground truth label for each image. When training the networks, the Mean Squared Error (MSE) between the predicted affective states and annotated affective states is minimised, and Adam Stochastic Gradient Descent is used to optimise the loss function (MSE), which is a fast optimisation algorithm for deep neural networks. Concordance Correlation Coefficient (CCC) [300] is employed to evaluate the performance of the models. CCC measures the correlation and degree of agreement between two variables. CCC has been widely used to evaluate models that predict continuous response variables [193, 194, 196, 198]. CCC ranges from -1 to 1, with perfect concordance at 1 and perfect discordance at -1.

CCC is calculated as follows:

$$CCC = (2\rho\sigma_x\sigma_y)/(\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2) \tag{3.4}$$

where μ_x and μ_y are the means for the two variables and σ_x^2 and σ_y^2 are their corresponding variances. ρ is the correlation coefficient between the two variables.

To train and evaluate the federated framework, data of 3 participants are randomly selected for evaluating the global model and the remaining 20 participants' data used for training 20 different models in a distributed manner. For the non-federated approaches (i.e., centralised processing of AUs and facial images), 8-fold cross validation is employed. During each training process, 7 folds are used for training and 1 fold kept aside for evaluation. The models are optimised by evaluating a range of hyper-parameter values. For instance, learning rates of 0.01, 0.001 and 0.0001; hidden sizes of 8, 16, 64, 128, 256 and 512; AU sequence lengths of 50, 100, 200, 400, 600, 800, 1000 and 2000; image sequence lengths of 4, 8, 16 and 32; and number of stacked recurrent layers of 1, 2, 4, 6 and 8. Table 3.6 presents the optimal hyper-parameter configurations of the models. The experiments for federated learning were split into 4 machines. One machine for aggregating the model weights obtained from the remaining three i.e., the global model. All machines consisted of a Graphics Processing Unit (GPU), 4 CPU cores and 6GB RAM. Our code is implemented in Pytorch with an epoch size of 100 for each experiment.

3.5 Results and Discussion

This section discusses the results of our methods on the different datasets and discusses the comparisons of our methods with other popular AI methods employed in the literature for predicting the driving behaviours.

Table 3.6: Hyper-parameter configuration of recurrent networks obtained using the RECOLA video and action unit datasets

Processing Method	Networks	Learning rate	Sequence length	Hidden size	Number of layers
Non-federated facial images	BiGRU	0.0001	16	8	1
	BiLSTM	0.0001	16	8	1
Non-federated action units	BiGRU	0.0001	600	512	6
	BiLSTM	0.0001	600	128	6
Federated facial images	BiGRU	0.0001	8	128	6
	BiLSTM	0.0001	8	128	6

3.5.1 Results of predicting driving styles

(a) Driving incident patterns and driving styles

After applying the 2-stage clustering framework and profile labelling algorithm on the unlabelled HGV telematics datasets, 11 HGV driving incident patterns (clusters) are identified as shown in Table 3.7. The labels ('low', 'moderate' and 'high') represent the occurrence of driving incidents in the different driving patterns obtained by the profile labelling algorithm.

Table 3.7: HGV driving clusters uncovered by the 2-stage clustering framework and profile labelling algorithm applied on the three driving incident datasets provided by Microlise: S for short, M for medium and L for long average mileage drivers.

Driving clusters	Harsh Braking	Over-Speed duration	Excessive Throttle	Number of Over Revs	Datasets
1	Low	Low	Low	Moderate	S,M,L
2	Low	Low	High	Low	M
3	Low	Low	High	Moderate	S
4	Moderate	Moderate	High	Low	L
5	Moderate	Moderate	High	Moderate	S,M,L
6	Moderate	Moderate	High	High	M
7	Moderate	High	Moderate	Low	S
8	Moderate	High	High	Low	L
9	High	Moderate	Moderate	Low	M
10	High	High	Moderate	Low	S,M,L
11	High	High	High	Low	M,L

Similar driving clusters are merged to produce the core HGV driving styles. Clusters 2 and 3 are merged as they differ only in number of over revs, one being low and the other being moderate. Similarly, clusters 4, 7 and 9 are combined with clusters 5, 8 and 10, respectively. While clusters 1, 6 and 11 are standalone driving styles as they do not follow the merging rule. Table 3.8 shows the seven core driving styles deduced from the 11 clusters uncovered across the four years. The first driving style represents a 'very calm' driving style with 'low' occurrence of driving incidents except number of over revs with 'moderate' occurrence. The second driving style is similar to the first except for a 'high' occurrence of excessive throttling, and it is labelled as 'calm' driving style. Driving style 3 is considered 'moderate' with 'moderate' harsh braking and over-speeding incidents. Driving style 4 is more aggressive than 3 with 'high' over revving incidents. Driving style 5 is considered 'speedy' with ‘high’ over-speeding and excessive throttling incidents. Driving clusters 9 and 10 produce an ‘aggressive’ driving style and cluster 11 produces a ‘very aggressive’ driving style with ‘high’ harsh braking, over-speeding and excess throttling incidents.

Table 3.8: Core Driving Profiles

Driving Styles	Cluster Numbers	N. of Harsh Braking Events	Duration of Over-speeding	Duration of Excessive Throttling	N. of Over Revs Events
Very Calm	1	Low	Low	Low	Moderate
Calm	2,3	Low	Low	High	Low-Moderate
Moderate	4,5	Moderate	Moderate	High	Low-Moderate
Moderate-Aggressive	6	Moderate	Moderate	High	High
Speedy	7,8	Moderate	High	Moderate-High	Low
Aggressive	9,10	High	Moderate-High	Moderate	Low
Very Aggressive	11	High	High	High	Low

(b) Driving style interpretability

The fuzzy logic stage of our framework models uncertainty in sensor data and imprecision in driving style definitions by representing both driving incidents and driving styles as FSs. In addition, the FLS produces interpretable driving styles by generating human-understandable mappings between the occurrence of driving incidents and driving styles. For example, instead of using precise number of driving incidents that is difficult to contextualise, the FLS transforms the number of driving incidents into ‘low’, ‘moderate’ and ‘high’ occurrence of driving incidents, which is easier for end-users and decision makers to understand.

The mapping of the occurrence of driving incidents to driving styles using IF-THEN rules represents cause-and-effect relationships. For example, Table 3.9 shows cause-and-effect relationships between the driving incidents and driving styles of short mileage HGV drivers. These relationships are important for educating drivers about the consequences of their driving and for road safety specialists and transport managers to revise current traffic laws according to the rules on speedy, aggressive and very aggressive driving styles. Subsequently, the output fuzzy sets and their membership degrees represent the uncertainties in predicting the different driving styles i.e., uncertainty in sensor readings, imprecision in human language, and variability in human interpretation.

(c) Comparison with conventional supervised learning approaches

The distribution of drivers in the different driving styles across the labelled datasets produced by the 2-stage clustering framework and profile labelling algorithms for the short, medium and long average mileage drivers is presented in Table 3.10. The datasets are split into training and test sets with a ratio of 70:30, respectively. The training training datasets are used to train RF, SVM, and MLP classifiers, as well as develop the fuzzy sets and rules of our fuzzy inference system. The datasets are imbalanced with majority of drivers having ‘Very Calm’ driving styles. As a result, Synthetic Minority Oversampling Technique (SMOTE) [301] coupled with random under-sampling of the majority class is employed to reduce class imbalance.

Table 3.9: Fuzzy rules for short mileage HGV drivers

Rule number	**Harsh braking**	**Over speeding**	**Excessive throttling**	**Over revving**	**Driving styles**
1	LOW	AND (LOW OR MODERATE)	AND (LOW OR MODERATE)	AND (LOW OR MODERATE)	Very Calm
2	MODERATE	AND LOW	AND (LOW OR MODERATE)	AND (LOW OR MODERATE)	Very Calm
3	MODERATE	AND MODERATE	AND (LOW OR MODERATE)	AND MODERATE	Moderate
4	HIGH	AND LOW	AND (LOW OR MODERATE)	AND (LOW OR MODERATE)	Aggressive
5	HIGH	AND MODERATE	AND LOW	AND MODERATE	Aggressive
6	LOW	AND MODERATE	AND HIGH	AND LOW	Calm
7	(MODERATE)	AND LOW	AND HIGH	AND MODERATE	Calm
8	LOW	AND (LOW OR MODERATE)	AND HIGH	AND MODERATE	Moderate
9	MODERATE	AND MODERATE	AND HIGH	AND (LOW OR MODERATE)	Moderate
10	(LOW OR MODERATE)	AND LOW	AND HIGH	AND LOW	Moderate
11	HIGH	AND MODERATE	AND HIGH	AND LOW	Aggressive
12	LOW	AND HIGH	AND MODERATE	AND LOW	Speedy
13	LOW	AND HIGH	AND HIGH	AND MODERATE	Speedy
14	LOW	AND HIGH	AND LOW	AND LOW	Speedy
15	MODERATE	AND MODERATE	AND (LOW OR MODERATE)	AND LOW	Moderate
16	(HIGH)	AND HIGH	AND (LOW OR MODERATE)	AND (LOW OR MODERATE)	Aggressive
17	HIGH	AND MODERATE	AND (LOW OR MODERATE)	AND LOW	Aggressive
18	HIGH	AND MODERATE	AND (MODERATE OR HIGH)	AND MODERATE	Very Aggressive

Table 3.10: Distribution of drivers among the different driving styles in Microlise's short, medium and long mileage datasets

HGV Driving Styles	**Short Average Mileage Dataset**	**Medium Average Mileage Dataset**	**Long Average Mileage Dataset**
Very Calm	1,692	2,387	2,601
Calm	173	43	-
Moderate	50	167	287
Moderate-Aggressive	-	459	-
Speedy	37	-	49
Aggressive	301	668	609
Very Aggressive	-	52	70
Total	2,253	3,776	3,616

Table 3.11 presents the average classification accuracy and f1-score of the supervised learning models on the test sets with comparison to our fuzzy inference approach. It is important to note that f1-score is a better evaluation metric compared to accuracy for evaluating the performance of the imbalanced datasets. The hybrid fuzzy approach shows best performance in the Short mileage dataset for accuracy evaluation metric and best performance in Medium and Long mileage datasets for f1-score. This means without defuzzifying the outputs of our fuzzy approach, the accuracy and f1-score of the approach may outperform the classifiers for the other datasets with lower performance. As defuzzification loses information about the uncertainties in predicting the driving styles. Other benefits of our framework compared to the classifiers are: (1) interpretation of HGV driving styles in terms of user-friendly descriptions of the occurrence of driving incidents (i.e., 'low', 'moderate' and 'high') and IF-THEN rules; and (2) representation of the uncertainties of predicting driving styles in terms of output fuzzy sets i.e., the likelihood of driving patterns belonging to driving styles.

In the future, class weight and focal loss functions could be explored to improve performance by forcing the models to focus more on minority driving styles (e.g., speedy and very aggressive driving styles) and difficult to classify driving patterns. In addition, telematics data with more valuable driving incidents could produce more reliable road safety HGV driving styles, such as harsh cornering, harsh lane changing and close following.

Table 3.11: HGV driving style classification results of popular conventional supervised learning methods and our fuzzy logic approach

	Accuracy (%)			**f1-score (%)**		
Models	**Short**	**Medium**	**Long**	**Short**	**Medium**	**Long**
RF	81.63	57.54	78.19	**77.69**	52.02	72.62
SVM	75.73	**70.24**	78.10	68.06	52.56	68.56
MLP	81.19	65.99	**85.48**	75.00	51.91	74.22
FLS	**84.44**	61.71	78.55	73.33	**60.00**	**76.00**

Table 3.12: The average accuracy (AA), average f1-score (AF), and average inference time (AIT) in seconds on 100 images and the average accuracy of each posture type after five test runs performed by models on the AUC 'split-by-driver' Distracted Driver test dataset

	VGG-19	**Densenet-201**	**Resnet50**	**InceptionV3-**	**InceptionV3-LSTM**	**InceptionV3-GRU**	**InceptionV3-BiGRU**	**InceptionV3-BiLSTM**
AA	0.833 ± 0.06	0.890 ± 0.03	0.877 ± 0.02	0.884 ± 0.01	0.902 ± 0.01	0.903 ± 0.01	0.917 ± 0.01	**0.917 ± 0.003**
AF	0.835 ± 0.06	0.895 ± 0.02	0.882 ± 0.02	0.890 ± 0.00	0.906 ± 0.02	0.909 ± 0.01	0.922 ± 0.00	**0.931 ± 0.004**
AIT	11.8 ± 0.34	16.6 ± 2.65	**11.6 ± 0.55**	12.3 ± 0.24	14.7 ± 0.29	13.9 ± 1.16	15.4 ± 1.77	16.1 ± 0.25
C0	0.86 ± 0.11	0.89 ± 0.05	0.89 ± 0.06	0.87 ± 0.08	0.93 ± 0.03	0.91 ± 0.05	0.95 ± 0.01	**0.95 ± 0.00**
C1	**0.97 ± 0.01**	0.83 ± 0.06	0.89 ± 0.13	0.83 ± 0.09	0.89 ± 0.03	0.87 ± 0.06	0.90 ± 0.08	0.82 ± 0.05
C2	0.99 ± 0.01	**1.00 ± 0.00**	1.00 ± 0.00	**1.00 ± 0.00**	0.99 ± 0.00	**1.00 ± 0.00**	**1.00 ± 0.00**	0.98 ± 0.01
C3	0.76 ± 0.20	0.96 ± 0.01	0.92 ± 0.08	0.97 ± 0.04	0.96 ± 0.03	0.98 ± 0.01	0.99 ± 0.03	**1.00 ± 0.00**
C4	0.94 ± 0.05	0.99 ± 0.01	0.98 ± 0.01	0.97 ± 0.03	0.98 ± 0.01	**0.99 ± 0.00**	0.98 ± 0.01	0.98 ± 0.02
C5	0.99 ± 0.02	**1.00 ± 0.00**	**1.00 ± 0.00**	0.99 ± 0.00	1.00 ± 0.00	0.96 ± 0.07	**1.00 ± 0.00**	**1.00 ± 0.00**
C6	0.82 ± 0.20	0.98 ± 0.06	0.98 ± 0.02	0.95 ± 0.04	0.90 ± 0.09	**0.98 ± 0.01**	**0.98 ± 0.01**	0.95 ± 0.01
C7	0.59 ± 0.21	0.59 ± 0.16	0.61 ± 0.09	0.62 ± 0.07	0.57 ± 0.08	0.59 ± 0.05	0.66 ± 0.17	**0.73 ± 0.13**
C8	0.62 ± 0.17	0.81 ± 0.13	0.69 ± 0.14	0.79 ± 0.09	**0.86 ± 0.14**	0.78 ± 0.08	0.78 ± 0.10	0.73 ± 0.08
C9	0.85 ± 0.10	0.93 ± 0.03	0.87 ± 0.06	0.95 ± 0.02	0.93 ± 0.05	0.96 ± 0.01	0.92 ± 0.09	**0.97 ± 0.02**

3.5.2 Results of predicting driving postures

To validate our hypothesis that recurrent neural networks can capture more discriminative temporal information to improve the prediction of driving postures (especially for postures with similar spatial representations), the results obtained from the proposed hybrid framework is compared with four state-of-the-art CNN architectures that only capture spatial discriminative information i.e. VGG-19, Densenet-201, Resnet50, and InceptionV3. The results are also compared with studies in the literature that explored driving posture detection using deep learning architectures and the AUC datasets.

(a) Comparison with state-of-the-art CNN models

Table 3.12 presents the average accuracy (AA), average f1-score (AF), average inference time (AIT) in seconds on 100 images, and the average accuracy of each posture on the AUC 'split-by-driver' distracted driver test dataset. First, it is observed that models implemented using our proposed framework, i.e. CNNs coupled with recurrent networks, outperformed the state-of-the-art CNN models in both AA and AF. This is due to the additional discriminative information between the feature maps captured by the recurrent networks. Amongst the models implemented using our proposed framework, InceptionV3-BiLSTM showed best performance with an AA of 0.917 (standard deviation 0.003) and AF of 0.931 (standard deviation 0.004). While Densenet-201 showed best performance amongst the CNN models with an AA of 0.890 (standard deviation 0.03) and AF of 0.895 (standard deviation 0.02). Statistically, the difference in accuracy between the InceptionV3-BiLSTM and Densenet-201 models is not significant with a p-value of 0.0802 (significance level is 0.05). However, the difference in f1-score (which is a more reliable evaluation metric than accuracy due for the imbalanced AUC training dataset) between the InceptionV3-BiLSTM and Densenet-201 models is statistically significant ($p = 0.0043$). In terms of model latency, ResNet50 outperforms the other models due to its shallow structure (fewer number of layers and connections between layers) compared to the other CNN models with a AIT of 11.6. The results show a trade-off between accuracy and model latency as deeper networks turn to produce more accurate results but take a longer time for inference e.g., DenseNet and the hybrid networks produce better performance but higher inference time. DenseNet is a very deep CNN model compared to the other CNN models with dense connections from all previous layers to all subsequent layers i.e. each layer receives the feature maps of all preceding layers. While the bi-directional hybrid networks are more complex compared to the other models because of the forward and backward data processing in addition to the convolution computation.

Furthermore, it is observed that "reaching behind" posture (c7) has the lowest average accuracy for all the CNN models compared to the other postures. This may be due to its spatial similarity with "talking to passenger" posture (c9) as shown in Figure 3.3. The CNN models appear to mistake "reaching behind" for "talking to passenger" postures. Our hybrid framework (i.e., InceptionV3-BiLSTM), however, improves the detection accuracy of "reaching behind" posture from 62% to 73%. Therefore, for the AUC database, extracting the spatial discriminative information from images and sequential discriminative information from the feature maps helps to better identify driving postures compared to using only spatial discriminative information obtained by CNN models.

Table 3.13: Driving posture classification results for studies that explore the AUC distracted driver test dataset

Model	Loss (NLL)	Accuracy (%)
VGG-16 [172]	1.2466	76.13
Resnet50 [172]	0.6615	81.69
Ensemble of AlexNet InceptionV3, ResNet50 and VGG-16 [172]	0.6400	90.06
Ensemble of ResNet50 and VGG16 [282]	0.3700	92.00*
InceptionV3-BiLSTM	**0.2793**	**91.70****

* has f1-score of 92.0% and ** has f1-score of 93.1%. Our hybrid framework results represented in bold.

(b) Comparison with other studies

To provide a comprehensive evaluation of performance, the framework is compared with the best results of CNN models in the literature that used the AUC distracted posture dataset [172, 282]. Table 3.13 presents the accuracy of the studies with comparison to our best performing hybrid model i.e., InceptionV3 coupled with BiLSTMs. Our model shows higher accuracy compared to individual CNN models employed in the literature but comparable performance with ensemble strategies of CNN models. This provides an opportunity for future work to take advantage of the strengths of individual CNN and RNN models by adapting our framework to consider an ensemble of CNN models and RNN models.

Furthermore, our proposed framework is limited to classification tasks only. That is, it cannot identify new types of distraction postures that are not found in the dataset it was trained on. For example, looking at outside objects or events has shown to be a dangerous distraction posture [302] but it is not found in the AUC dataset. Therefore, our framework will not be able to detect this new distraction posture. The framework is also limited to visual and manual distractions. Cognitive distractions where a driver is lost in thought e.g. daydreaming, hand-free calling and listening to the radio, cannot be recognised using our framework. For future work, to make our framework more robust to new distraction types and cognitive distractions, more data that capture those distractions are needed. In addition, the comprehensive data can be explored using anomaly detection techniques to model and distinguish between "safe driving" and "distracted driving".

3.5.3 Results of predicting affective states

This section presents the results of our proposed federated deep learning framework with comparison with non-federated deep learning strategies applied to the RECOLA database.

(a) Federated vs non-federated strategies

The performance of federated facial affect recognition is compared with non-federated strategies using CCC evaluation metric. Table 3.14 shows the CCC for predicting valence and arousal affective states. The bold values represent the best model performance. Overall, the non-federated processing of facial images shows best valence and arousal predictions, followed by the federated processing of facial images. The strategies that process the raw facial images outperform the processing of AUs due to the loss of spatial information in the AUs. CNNs coupled with BiLSTMs show best performance for non-federated processing of images. The processing of AUs shows similar arousal prediction performance compared to the federated processing of images. In addition, It can be observed that LSTMs outperform GRUs when processing the images similar to results obtained from driving posture detection (Section 3.5.2). However, for AU processing, GRUs show better performance compared to LSTMs. This is due to the efficiency of GRUs in processing smaller datasets or feature sets compared to LSTMs as only 40 facial features are extracted by the facial landmark extractor while 512 non-linear features are extracted by the convolutional networks.

Table 3.15 presents the efficiency results of the best performing models in terms of data protection, training time and inference time. Processing AUs has the least training and inference times due to a smaller feature set (which reduces the complexity of the network) and lack of the convolutional feature extraction layer. This makes the AU processing modules more suitable for real-time affect recognition systems such as, real-time monitoring of drivers' affective states for early intervention and assistance. However, the predictive accuracy of processing AUs is lower compared to the other strategies. The non-federated processing of images shows better accuracy in predicting valence and arousal compared to AUs and FL at the detriment of the potential exposure of users' facial identities. FL best preserves users' identities and sensitive information compared to the other methods as data is maintained in users' local machines, however, its training time is significantly higher, which can further increase if the processing at the local machines is not done synchronously. Lastly, FL's CCC results are inferior to the non-federated processing of images due to limited data at the local machines.

It is important to mention that the performance of the models are based on the averaged valence and arousal ratings provided by the six annotators. In the future, the ratings of the individual annotators or other aggregation methods (such as majority voting for continuous values [293]) could be explored to improve the reliability of the ground truth labels, as the reliability of model performance depends highly on the reliability of the response surface.

Table 3.14: Average CCC for predicting valence and arousal using variations of RNN models on RECOLA datasets (best performance in bold).

Affect Recognition Method	**Deep Learning Architecture**	**Valence CCC**	**Arousal CCC**
Non-federated Processing of Facial Images	CNN-BiGRU	0.415	0.504
	CNN-BiLSTM	**0.476**	**0.515**
Non-federated Processing of Action Units	BiGRU	0.347	0.401
	BiLSTM	0.269	0.365
Federated Processing Facial Images	CNN-BiGRU	0.393	0.273
	CNN-BiLSTM	0.426	0.390

Table 3.15: Model performance, training time and inference time for the best performance deep learning strategies explored for affect recognition (best performance in bold).

Affect recognition methods	**Level of Protection**	**Valence CCC**	**Arousal CCC**	**Training time (minutes)**	**Inference 100 images (seconds)**	**Inference 500 images (seconds)**
Non-federated facial images	Exposed	**0.476**	**0.516**	315.4	17.509	22.921
Non-federated action units	Partial Protection	0.347	0.401	**141.5**	**1.351**	**2.962**
Federated facial images	**Protected**	0.426	0.390	599.2	19.720	26.860

Table 3.16: Comparison of valence and arousal predictions between our proposed methods and other studies using RECOLA datasets (best performance in bold).

Affect Recognition Method	**Type of Machine Learning Model**	**Valence CCC**	**Arousal CCC**
Non-federated Processing of Facial Images	CNN + LSTM [193]	**0.620**	0.435
	DNN [196]	0.379	0.464
	CNN + LSTM [198]	0.538	0.336
	CNN + RNN [195]	0.474	-
	2D/3D CNN + ConvLSTM [194]	0.546	-
	Our CNN + BiLSTM	0.476	**0.514**
Non-federated Processing of Action Units or Facial Landmarks	LSTM [198]	0.483	0.137
	SVM [197]	**0.507**	0.272
	BiLSTM + SVM [199]	0.394	0.265
	Our BiGRU	0.347	**0.401**
Federated Processing of Facial Images	**Our CNN + BiLSTM**	**0.426**	**0.390**

Note: A dash is inserted if the results were not reported in the original papers.

(b) Comparison with other studies

Table 3.16 compares the performance of our models with other studies that employ deep learning methods on the RECOLA image and AU datasets for affect recognition. For the non-federated processing of facial images, it is observed that [193, 194, 196, 198] show better valence recognition results compared to our non-federated deep learning model, with Tzirakis *et al.* [193] having the best CCC valence (*0.620*).

However, our model shows best arousal accuracy with a CCC value of *0.514*. Those studies also explored different architectures of CNNs coupled with LSTMs, but feed the outputs from fully connected layers in the CNN stage into the LSTMs. These outputs from their fully connected CNN layers are complex non-linear representations of the relationships between the spatial features (pixels) of the images. In addition, the studies adopt a train-test split evaluation approach, which does not provide a comprehensive exploration and evaluation of the data.

Furthermore, the processing of AUs and facial landmarks by previous studies [197, 198, 199] show better CCC results in predicting the valence dimension. However, our AU deep learning model outperforms the arousal accuracy of the other studies (*0.401*). This is due to the remarkable performance of GRUs in processing small feature sets. Storing and processing the anonymised facial features (AUs) is more secured in terms of privacy compared to storing and processing the raw facial images. In order to maintain and process a database of facial images, appropriate security levels and systems to safeguard the data are required. In addition, our review of AUs for affect recognition in Section 3.2.3 revealed that AUs are not completely secured as image reconstruction techniques (e.g., auto-encoders and generative adversarial networks) could potentially reconstruct faces from facial landmarks. Therefore, our proposed federated learning strategy protects users' data compared to the centralised processing of AUs. We could not find any study in the literature that explores FL for facial affect recognition to compare with our federated results of *0.426* for valence and *0.390* for arousal.

Furthermore, the results of federated vs non-federated strategies show a trade-off between efficiency and privacy for facial affect recognition. Consequently, from a privacy-compliant and data protection perspective, it could be argued that storing facial images may not be necessary if other alternative methods that produce acceptable performance are available. However, our best result for federated processing of facial images(i.e., valence CCC equals *0.426* and arousal CCC equals *0.390*) is not sufficient for deployment as 0.426 is still far from perfect agreement i.e., CCC of 1. Therefore, for driver affect recognition, highly secured non-federated facial affect recognition models are still the most accurate to deploy in driving assessment systems. In addition, more work needs to be done to improve the performance of federated affect recognition, by analysing additional data sources (such as voice patterns and eye movements), and by exploring attention mechanisms that focus more on parts of the image relating to human emotions.

Lastly, in the future it will be important to also consider the following characteristics of human emotions when improving the accuracy and reliability of driver affect recognition systems: (1) differences between facial expressions and what people actually feel; and (2) cultural differences in expressing and reading emotions.

3.5.4 Implications of the results to HGV driving assessment

The above sections have presented the results of our proposed methods in predicting the different facets of driver behaviour that impact driving risk from driver data i.e., drivers' driving styles, driving postures and affective states. For an online data-driven assessment of HGV driving risk, these methods would process HGV driver data streams to accurately characterise the driving behaviours, and forward the information to an intelligent risk assessment component. The intelligent driving risk assessment component analyses the information to compute the impact of the driving behaviours on road safety. However, the AUC and RECOLA databases utilised for evaluating the hybrid deep learning framework introduced in this chapter (for detecting driving postures and affective states) were collected from non-HGV drivers in controlled settings. This was due to the challenges of obtaining naturalistic HGV driver footage, such as obtaining drivers' consent to use their data, developing secure storage systems to protect the data, and developing privacy compliant procedures and technologies to process the data. Nevertheless, this does not undermine the application of our methods to HGV road safety driving assessment. For a more reliable identification of HGV driving postures and affective states in the future, our methods should be retrained and evaluated on well labelled HGV driver footage as HGV drivers may display different driving postures and affective states compared to non-HGV drivers.

Another limitation of the data is that the different databases were not collected together i.e., the footage data is not affiliated with the telematics data. As a result, it is difficult to determine the relationships among drivers' driving styles, postures and affective states using those datasets. This also does not allow us to simultaneously detect the different driving behaviours and provide more reliable results. However, our methods have shown state-of-the-art performance in detecting driving styles, postures and affective states, and therefore, can be easily deployed into systems that capture HGV telematics data and driver footage for further improvement of HGV driver behaviour characterisation. Chapter 5 captures the relationships among driving styles, postures and affective states by engaging with key stakeholders in the HGV sector, who possess qualitative insights about the driving behaviours.

As revealed by the psychological theories on driver behaviour in Chapter 2, the different facets of driver behaviour are continuously influenced by external conditions to impact driving risk, such as weather conditions, road types and traffic conditions. Therefore, in order to reliably assess the risk of driving behaviours, it is important to incorporate the impact of inevitable external conditions.

3.6 Summary

This chapter introduced a driver behaviour characterisation framework made up of three intelligent methods to improve the prediction accuracy, interpretability and data privacy of driving styles, driving postures and affective states. First, it presented a novel hybrid fuzzy logic framework to automatically identify and interpret core driving styles from unlabelled telematics data. The framework was employed on a large telematics driving incident data captured from more than 2,000 HGVs in the UK from 2014 to 2017. The results demonstrate better performance than conventional machine learning approaches in modelling data uncertainties, and producing more robust and understandable descriptions of driving styles. Secondly, it presented a novel hybrid deep learning framework for detecting driver distraction postures in driver footage. The framework consists of stacked CNNs to extract feature maps from the images that represent the spatial discriminative features between the postures, and stacked recurrent networks to capture the sequential discriminative relationships among the spatial features. The framework is applied on the AUC distracted driver dataset and compared to other popular CNN architectures such as VGG, Inception, ResNet and DenseNet. The framework outperformed the CNN models in both average accuracy (91.7%) and average f1-score (93.1%). Lastly, it presented a federated implementation of deep learning methods for affect recognition using facial images while protecting users' identities and privacy. The federated approach is compared to non-federated strategies i.e., centralised storage and processing of facial images, and anonymised facial features. The results of the strategies when applied to the RECOLA image and action unit datasets show trade-offs between accuracy, efficiency and privacy. The centralised processing is less secured compared to the federated framework, but more accurate in predicting valence and arousal. However, the results show the potential of detecting valence and arousal in an online system, where data streams are processed in local devices to protect the identities of drivers.

In the next chapter, the contextual characteristics of real-world driving environments are considered, such as weather conditions, traffic conditions, road types and work schedules, in the assessment of HGV driving performance and risk. The impact of these factors on drivers' responses are commonly not captured in driver data [9, 303], but at least can be partially obtained from stakeholders' experiences in the HGV sector. Therefore, simply assessing drivers' behaviours and not the environment in which they operate is not sufficient to provide a realistic, reliable, fair and holistic assessment of driving risk. The chapter introduces a novel systematic stakeholder-supported approach to capture and embed the impact of external conditions into the assessment of HGV driving behaviours when comprehensive driving risk data are scarce or unavailable.

Chapter 4

Modelling the Impact of Contextual Factors

4.1 Introduction

The previous chapter introduced an intelligent multifaceted driver characterisation framework to predict driving behaviours using three data sources obtained from sensors. However, as revealed by the psychological theories on driver behaviour in Chapter 2, the multifaceted HGV driving task is mostly influenced by external contextual factors, such as, weather, traffic, road conditions, time pressure for delivery and work schedules. Therefore, in order to fully understand and assess the impact of predicted HGV driving behaviours on road safety, the impact of external contextual factors on drivers' responses need to be considered.

This chapter answers the third research question of this thesis: "how can we reliably capture, model and incorporate information about the impact of contextual factors on drivers' responses and road safety into the assessment of HGV driving risk?". It introduces a novel framework called stakeholder-supported intelligent driving assessment (SIA) that engages with key stakeholders in the HGV sector to capture the influence of driver traits and external factors on HGV driving performance. Subsequently, the variability and imprecision in the opinions of the stakeholders about the impact of the factors are modelled using Fuzzy Sets (FSs) based on the Interval Agreement Approach (IAA) [213, 257, 304], fused and incorporated into the assessment of driving performance and risk.

The chapter is organised as follows. Section 4.2 provides an overview of current intelligent data-driven driving risk assessment (CIDA) approaches reviewed in Chapter 2 (page 28). Subsequently, the rationale of using the IAA approach to capture and model expert input is provided. Section 4.3 describes the different stages of SIA, and Section 4.4 presents the experimental design for the application of SIA in HGV driving risk assessment. The results of its application are discussed in Section 4.5, and lastly, Section 4.6 summarises the chapter.

4.2 Background

This section provides an overview of existing intelligent data-driven driving risk assessment approaches and the motivation for engaging with stakeholders to capture contextual information. It also describes the IAA approach, which is employed to capture and model the modulation of stakeholders views of perceived risk factors.

4.2.1 Related work

According to our review of CIDA approaches presented in Section 2.4.2 (page 28), researchers have mainly focused on behaviour-centric driving risk assessments. Those studies explored different computational and artificial intelligence techniques on driver data (e.g., driving incident data [14, 15] and GPS data [16, 17, 305]) that represent the manner by which drivers operate vehicle controls (i.e., driving styles) to determine the level of driving risk. Narrow subsets of external factors have also been explored together with driving styles [16, 229, 252, 305]. Analysis and assessing driving risk using solely drivers' driving styles could potentially lead to incomplete and unfair assessments as real-world HGV commercial driving is affected by inevitable contextual factors, such as individual drivers' physical and mental states, weather conditions, traffic conditions, road types, work schedules and time pressure for delivery. On the other hand, information about the impact of these inevitable contextual factors and driver behaviour on road safety is not available based on current data sources, but at least can be partially obtained from stakeholders in the industry who possess a deep understanding of the highly complex and dynamic environment.

As revealed in Section 2.4.1 (page 27), researchers in social sciences have engaged with stakeholders in the driving sector to understand the impact of external factors on road safety, focusing primarily on drivers [7, 49, 50, 306, 307, 308]. However, other crucial stakeholders exist in the sector who could potentially provide more insights, such as managers, road safety officers and researchers. In addition, the results from the studies are difficult to incorporate into online driving assessment systems.

SIA is developed to address the above shortcomings by engaging with drivers, managers, road safety officers and researchers to identify relevant contextual factors that affect HGV driving performance and capture their impact on road safety. The contextual information are later modelled using IAA FSs and fused to determine the impact of detected driving behaviours and perceived factors on road safety driving risk.

4.2.2 Insights from stakeholders

The literature shows that questionnaires have been the main tools to capture insights from human participants [232, 257, 308, 309, 310, 311, 312, 313]. Typically, the questionnaires are made up of discrete-valued response-format questions that only allow one point on the response scale to be selected by participants [308, 311, 312, 313]. Commonly used discrete ordinal scale response formats are a five-point or a seven-point scale from ‘strongly disagree’ to ‘strongly agree’. Data from point-based or discrete-valued response scales are relatively easy to analyse (e.g., by using barplots, pie charts and statistical techniques); however, they are not suitable for capturing uncertainty and imprecision in human perception and views [232].

As an alternative, interval-valued response-format questionnaires have been developed to capture and better quantify uncertainty in individual responses in several domains e.g., education [314], hospitality [315, 316], service quality [317, 318, 319], impact assessment [320] and customer preferences [321]. In HGV driving, this uncertainty may reflect limitations of stakeholders’ knowledge about complex relationships between contextual factors and driving risk. For example, the precise effect of weather conditions on road safety is difficult to determine due to interactive effects of other contextual factors, such as time pressure, road types and time of the day.

Furthermore, regardless of the response-format of questionnaires, the insights or responses obtained from human participants are likely to differ due to differences in precision, perception, experiences and expectations [232, 255, 322]. Thus, it is expected that different stakeholders —even though they have similar roles — may provide different answers to questions. In addition, stakeholders with different roles may have varying viewpoints due to their distinct responsibilities and contrasting interests [256]. Those input differences must be effectively captured and modelled to provide a comprehensive, reliable and clear representation of knowledge to support collaborative decision-making.

4.2.3 Rationale for using the Interval agreement approach

A modelling technique called the interval agreement approach (IAA) [257] was developed to model, combine and provide a clear representation of imprecision and uncertainties in human insights.

Compared to other approaches for capturing and modelling questionnaire data, the IAA has several advantages. The IAA is a more flexible approach than traditional quantitative methods, such as discrete scales. Traditional quantitative methods require participants to select a single response from a set of options, which can be limiting and may not accurately capture the uncertainties of their opinions.

The IAA is simpler to implement, understand and visualise compared to other approaches that allow imprecise and uncertain information (such as Fuzzy Delphi method [323] and Fuzzy Analytic Hierarchy Process [324]; as it does not require any group facilitators and weighting strategies, and it follows a least commitment principle [325].

4.2.4 Interval agreement approach

The IAA models agreement (or overlap) between the responses of participants captured using interval-valued surveys. For example, Figure 4.1 presents an IAA fuzzy set constructed from the union of three interval responses, with higher agreement in areas where the intervals overlap.

The IAA generates non-parametric Fuzzy Sets (FSs) capturing all different levels of uncertainty in individual opinions and also between multiple individuals/groups’

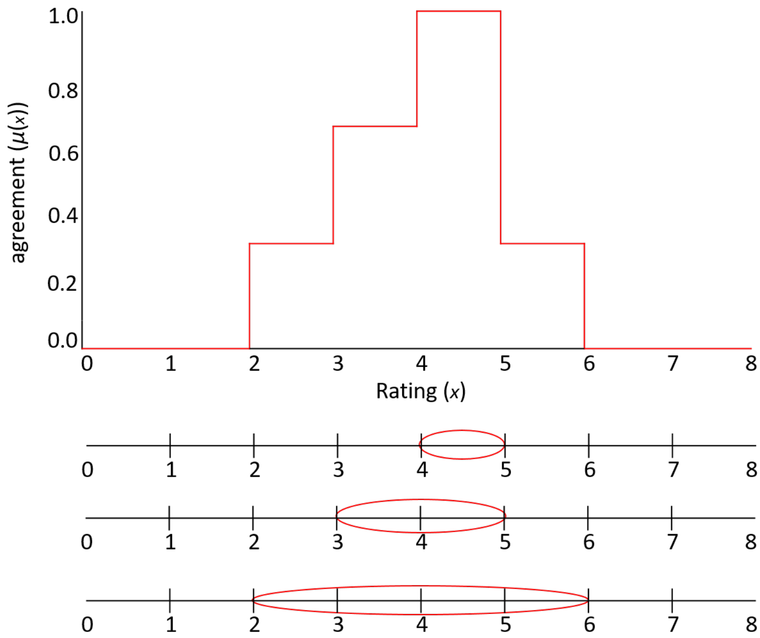


Figure 4.1: Illustrative example of how IAA captures the agreement between three intervals.

opinions. During IAA FSs model creation, the collected intervals are first formed in Type-1 FSs (T1 FSs), which minimises the loss of information in participants' opinion. Secondly, these generated T1 FSs are aggregated to z-slice representations of General Type-2 FSs (zGT2 FS) that allow to model different individual opinions from different groups of professions all together.

For example, in Table 4.1, experts 1-3 -who are working in the same profession A- are asked a question and they provide intervals which allow to capture uncertainty in their opinions. In IAA, these collected intervals from each individual are aggregated into a single T1 FS. Thus, the generated T1 FS is able to capture different opinions from experts (Expert 1-3) and model them in a single representation that shows the aggregated opinions of experts from the same profession A, as shown in Figure 4.2 (a). The y-axis ($\mu(x)$) represents the level of agreement among the responses e.g. the experts show greatest agreement in their responses at '2'.

Another group of experts (Expert 4-6) -who are working in profession B- are asked the same questions and they provide different opinions with different levels of uncertainty represented by intervals, as illustrated in Table 4.2. These intervals can be aggregated into another single T1 FSs, which is shown in Figure 4.2 (b). As can be seen in the comparison of Table 4.1 and 4.2, the experts 4-6 tend to be more uncertain about their opinions which leads to a *wider* T1 FS in Figure 4.2 (b).

Table 4.1: A sample of collected intervals from three experts (1-3) with the profession A

	Expert 1	**Expert 2**	**Expert 3**
Profession A	[1,2]	[1,3]	[2,4]

Table 4.2: A sample of collected intervals from three experts (4-6) with the profession B

	Expert 4	**Expert 5**	**Expert 6**
Profession *B*	[1,5]	[1.5,4]	[1,6]

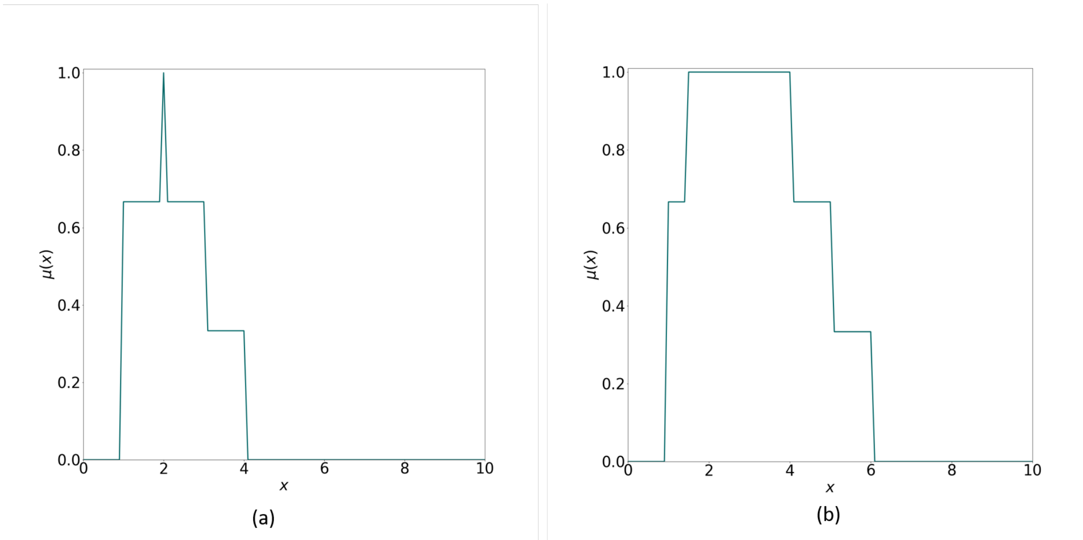


Figure 4.2: T1 FSs for each profession using the IAA approach. (a) IAA-Profession A, and (b) IAA-Profession B.

Subsequently, the T1 FSs are aggregated to generate zGT2 FSs where the agreement/variation between multiple experts/groups of information is modelled through the secondary memberships (z-slices), as demonstrated in Figure 4.3. The darker area (a) in the plot represents the region with higher agreement between the two IAA FSs i.e. both groups agree the most in their responses between 1 and 4.

T1 FSs are defined in Section 2.3.4 (equation 2.11). A GT2 FS [326] F is characterised by a MF $\mu_F(x,u)$, where x $\in X$ and u $\in J_x \subseteq [0,1]$, i.e

$$F = \{((x,u), \mu_F(x,u)) \mid \forall x \in X \; \forall u \in J_x \subseteq [0,1]\} \tag{4.1}$$

in which $\mu_F(x,u) \in [0,1]$. F can also be expressed as follows:

$$A = \int_{x \in X} \int_{u \in J_x} \mu_F(x,u)/(x,u), \;\; J_x \subseteq [0,1] \tag{4.2}$$

where $\int\int$ denotes the union over all admissible u and x.

A zGT2 FS is formed by slicing a GT2 FS in the third dimension (z) at level z_i. This slicing action will result in an interval set in the z-dimension with height z_i. Therefore, a zGT2 FS is a z-slice or interval representation of a GT2 FS, where the z-dimension is not fixed to 1 but is equal to z_i, where $0 \leq z_i \leq 1$.

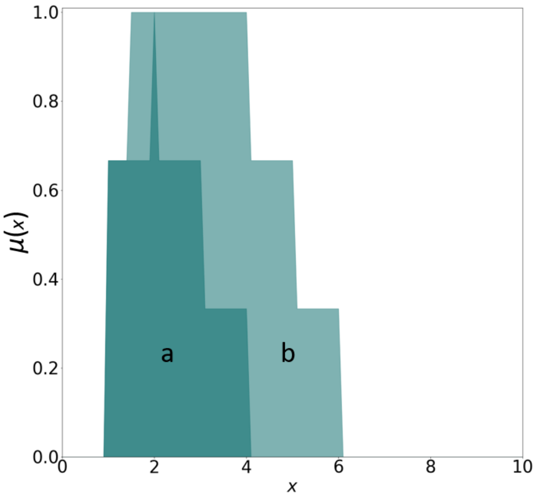


Figure 4.3: 2D view of the zGT2 FS produced with the IAA T1 FSs from Figure 4.2

4.3 Stakeholder-Supported Driving Assessment Framework

This section introduces our stakeholder-supported context-aware framework, SIA. SIA extends CIDA approaches by engaging with stakeholders to capture the impact of contextual factors on driving risk. The perceived contextual information is modelled using IAA FSs and integrated into the decision making process to ensure a collaborative, comprehensive, realistic and fairer assessment of driving risk.

Figure 4.4 shows a diagram with SIA's stages and the extension of CIDA methods.

4.3.1 Stage 1: Identification of contextual factors

This stage involves compiling a list of contextual factors that affect driving performance based on the literature. Subsequently, consultative workshops are organised with stakeholders in the industry who possess qualitative insights about the domain. Those include drivers, transport managers, road safety professionals, traffic officers and road safety researchers in the HGV sector. During the workshops, the factors from the literature are presented to the stakeholders. Stakeholders are invited to share their opinions on two questions:

1. "Are the contextual factors identified from the literature sensible and valid?"

2. "Are there any other contextual factors that should be considered?"

These workshops are conducted to validate and update the factors obtained from the literature, as some of those factors may be outdated due to advances in technologies and regulations. They also capture the factors that stakeholders

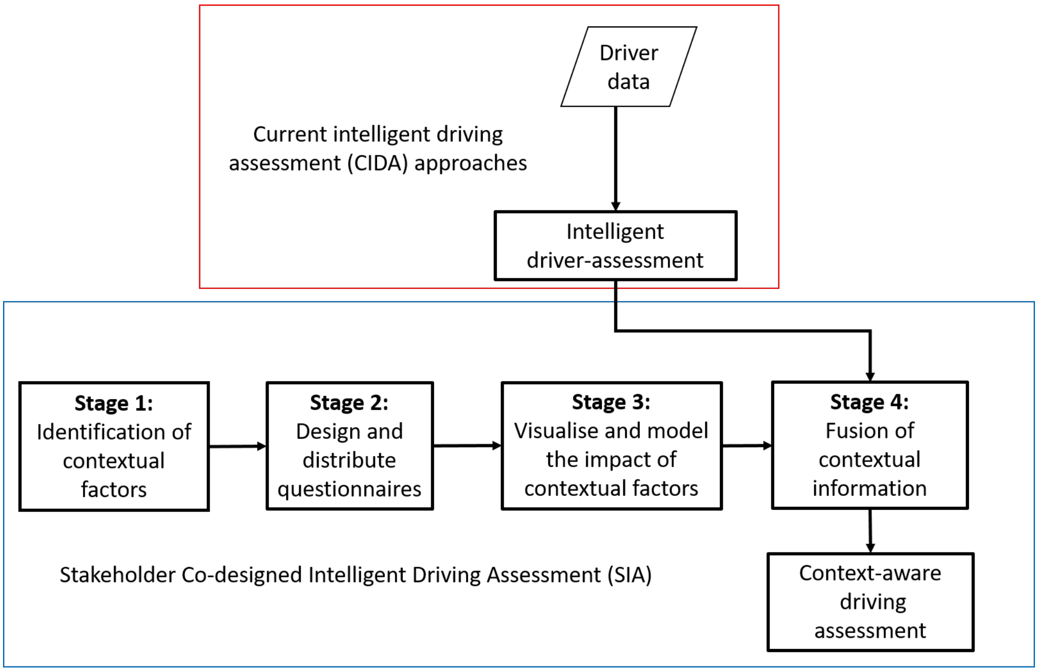


Figure 4.4: A diagram showing an extension of CIDA approaches (red bounding box) with a new stage (SIA) that captures and incorporates information about the impact of contextual factors on road safety.

consider when assessing safe driving to incorporate expert judgement into the driving risk system.

4.3.2 Stage 2: Design and distribution of questionnaires

After identifying and validating the factors in Stage 1, interval-valued response-format questionnaires [232] are designed to capture stakeholder insights about the impact of the factors as well as imprecision or indecision in their opinions. The design and wordings of the questions are improved during the workshops with the stakeholders recruited in Stage 1. The stakeholders are asked the following questions:

1. "Is the format and rating scale of the questionnaire easy to comprehend and complete?"

2. "Are the wordings and questions understandable?"

3. "Is the length of the questionnaire adequate considering the busy schedules of professionals?"

4. "Is there anything important missing in the questionnaire?"

The final questionnaire is distributed to a wide cohort of stakeholders to obtain their views about the impact of the contextual factors on driving risk.

4.3.3 Stage 3: Visualising and modelling the impact of contextual factors

In this stage, the responses from the questionnaires are gathered and transformed into an appropriate format for visualisation and analysis e.g., using tuples to represent the interval of responses. A least commitment strategy (i.e. no outlier removal) [325] is employed to analyse the data as outliers may contain rich information and IAA FSs are able to effectively handle this.

The transformed data are visualised using line graphs and box-plots to understand individual as well as group opinions, respectively. The data plots assist in understanding: 1) the difference in opinions across the different professions, 2) the level of certainty in the responses of stakeholders, and 3) the agreement amongst stakeholders within each profession.

The set-based mechanism of IAA is adopted to model the data using IAA FSs. The FSs account for any variability in the responses of stakeholders. Therefore, if there exist n contextual factors and i professions, $n \times i$ IAA FSs are generated. To understand and quantitatively express the level of agreement among the opinions of different stakeholder professions, the similarities between their fuzzy sets are measured. This is important for the development of a reliable and collaborative system to ensure that the system is not biased towards particular stakeholders.

4.3.4 Stage 4: Fusion of contextual information

The IAA FSs are aggregated into zGT2 FSs [257] by employing the agreement principle in Wagner *et al.* [327] and associating a higher secondary membership (zLevel) to areas where the IAA FSs overlap. That is, if $n \times i$ IAA FSs are generated for n contextual factors and i professions, each FS representing a specific factor is aggregated with their corresponding FSs to produce n zGT2 FSs. The secondary membership captures the agreement among the different professions and the zGT2 FSs provide the representation and separation of the individual types of uncertainty present in the data. The resulting zGT2 FSs produced in this stage represent the overall impact of the contextual factors with different levels of uncertainty.

To provide an easy integration solution for the contextual information (i.e., zGT2 FSs) into intelligent driving assessment systems, the zGT2 FSs are defuzzified and fused.

Section 4.5 presents the results of the application of SIA in capturing, understanding and fusing information about the impact of contextual factors on HGV driving risk in the UK, which will elucidate the above stages.

4.4 Experimental Design

HGV driving is mostly influenced by drivers' personal traits and external contextual factors. The influence of these factors are not captured in existing driver data and/or artificial intelligence models; however, they can at least partially be obtained from stakeholders expertise in the HGV sector. This section presents the experimental design of applying SIA to capture information about the impact of contextual factors on HGV drivers' responses from stakeholders in the HGV sector. The section also presents the data modelling and processing techniques based on fuzzy sets adopted in stages 3 and 4 of SIA to understand and incorporate the contextual information into online data-driven decision-support systems.

4.4.1 Workshops with stakeholders

To identify and validate the contextual factors that affect HGV driving risk, stages 1 and 2 of the framework are applied. Five iterative workshops were organised with nine stakeholders consisting of a university professor in Psychology specialised in HGV driver behaviour analysis, three HGV transport managers who supervise HGV drivers to optimise deliveries and ensure their companies comply with road safety regulations, and five researchers specialised in driving behaviour analysis and fuzzy logic. In each iteration, the stakeholders refined the factors, questionnaire design, format and instructions. The workshops were held virtually due to COVID-19 restrictions between September 2020 and November 2020. The collaboration between nine stakeholders were sufficient to identify, update and validate the factors and questionnaire design because too many stakeholders would have been difficult to manage. To complement the lack of HGV drivers who are the operators of the vehicles and road safety officers who enforce road safety regulations on road users in the workshops, the first two HGV drivers and road safety professionals who completed the questionnaire were interviewed, asking them if there were any other important factors missing from the questionnaire and whether the design of the questionnaire was appropriate. Their responses were used to update the questionnaire as reported in Section 4.5.1.

4.4.2 Questionnaire design

The final questionnaire consisted of nine-point scale questions that asked participants to provide their opinions or ratings about the impact of the factors on HGV driving performance. The nine-point rating scale ranged from 1, meaning 'strong negative impact', to 9, 'strong positive impact' and 5 representing 'no impact', as shown in Figure 4.5. The number of points in the scale were decided in the workshops as stakeholders found nine-point ratings more comprehensive to express their opinions. Stakeholders were more familiar with discrete-valued response-format scales compared to interval-valued scales. Therefore, in order to capture imprecision or uncertainty in the opinions of stakeholders and still obtain a sufficient number of responses, participants were instructed to select two discrete points representing the range of certainty of their responses. This approach is not the same as continuous interval responses, but it ensures easy design, administration and completion of the questionnaire. The final questionnaires are found in Appendix A.1.

4.4.3 Participant recruitment

Professionals from the four stakeholder groups presented in Figure 4.6 were recruited to complete the questionnaire. The figure provides summarised definitions of the

Considering the following external factors, how do you think they could affect a driver's performance?

	Strong Negative Impact (1)	2	Negative Impact (3)	4	No Impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
High traffic	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low traffic	☐	☐	☐	☐	☐	☐	☐	☐	☐
Time pressure for delivery	☐	☐	☐	☐	☐	☐	☐	☐	☐

Figure 4.5: Sample questions extracted from the questionnaire with nine-point rating scales to capture expert opinions about the impact of contextual factors on HGV driving performance.

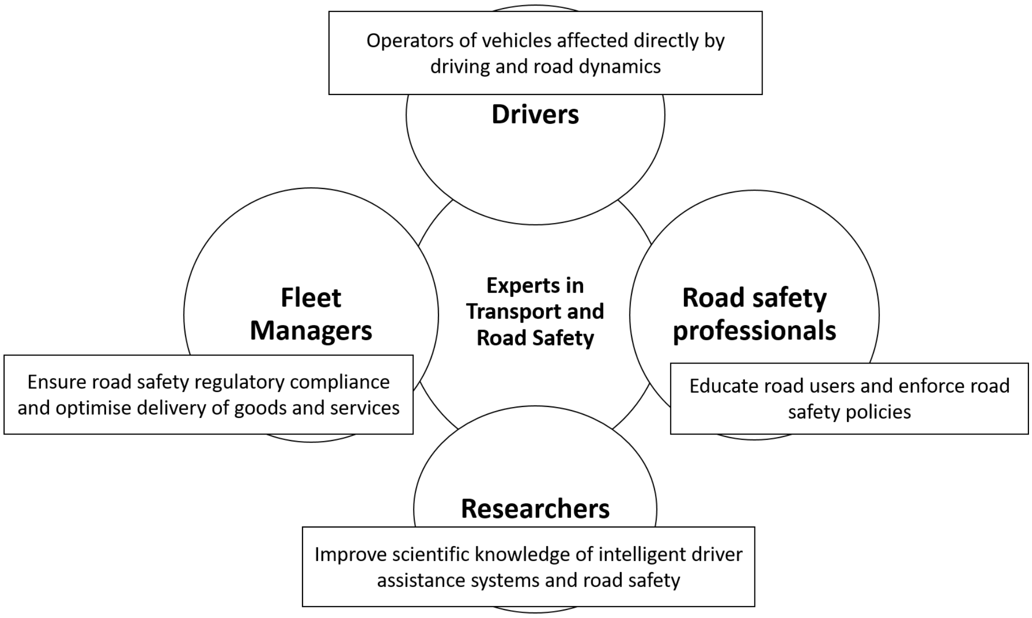


Figure 4.6: Diagram showing the different types of stakeholders in the HGV industry along with the definitions of their roles.

different stakeholder groups. Professionals were recruited by sending mass messages to individuals on LinkedIn [328], University of Nottingham, and UK's University Transport Study Group whose job titles and expertise matched any of the four stakeholder groups. No compensation was offered for participation. Table 4.3 provides a summary of the number of participants recruited and the average years of experience. Ninety-three participants from the UK completed the questionnaire. Among the participants were: 20 HGV drivers, 23 researchers, 24 HG managers, and 26 road safety professionals. Years of experience ranged from 3 to 46 years with average and standard deviations of (M=22.90, SD=12.79) for HGV drivers, (M=17.33, SD=10.72) for researchers, (M=11.04, SD=8.22) for HGV managers, and (M=20.81, SD=12.39) for road safety professionals.

Table 4.3: Distributions of participants, average years of experience, and standard deviations of experience among the groups of stakeholders

	HGV drivers	**Researchers**	**HGV managers**	**Road Safety Professionals**
No. of Participants	20	23	24	26
Avg. Experience (Yrs)	22.90	17.33	11.04	20.81
Std. Experience (Yrs)	12.79	10.72	8.22	12.39

4.4.4 Questionnaire data modelling techniques

The two-point discrete responses selected by participants are transformed to intervals to effectively represent uncertainty using the interval agreement approach described in Section 4.2.4. This captures set-valued information based on the 9-point scale. IAA FSs are generated from the two-point discrete responses and later aggregated into zGT2 FSs by weighting areas where the IAA FSs overlap. Jaccard similarity measure [329] is employed to calculate the agreement in opinions amongst the different groups of stakeholders. The Jaccard similarity measure is an efficient and well-established method used to calculate similarity between fuzzy sets. It calculates the cardinality of the intersection of two sets, divided by the cardinality of the union of the two sets. The output value for the method lies between 0 and 1, where 1 indicates total agreement and 0 indicates total disagreement. For easy integration of the contextual information (i.e., zGT2 FSs) into the assessment of HGV driving risk, the FSs are reduced to crisp values using centroid type-reducer (CTR) defuzzification [330] and employ product aggregation strategy to fuse information about perceived contextual factors. The CTR method takes into account the centroids of the primary, lower and upper membership functions, weighted by their respective membership values, to obtain the defuzzified output.

4.5 Analysis, Results and Discussion

This section presents the analysis of the responses provided by the stakeholders. In addition, it provides a solution to incorporate and fuse information about the impact of perceived contextual factors into the assessment of HGV driving performance and risk.

4.5.1 Contextual factors identified from the workshops

Figure 4.7 presents external factors and driver traits identified from the literature that impact HGV driving performance [6, 7, 12, 49, 50, 233, 234, 235, 237, 306, 307, 331, 332, 333, 334]. The factors were presented to stakeholders in the workshops and some were identified as irrelevant, such as vehicle characteristics. Others were identified as outdated due to new road safety policies in the UK, such as rest breaks. Time of the day and day of the week were revised to start, mid and end of shifts, as HGV drivers start their jobs at different times of the day and different days of the week. The stakeholders proposed additional factors, such as driver confidence, impact of different temperatures and road types. The figure shows the updated list of contextual factors identified by the stakeholders in the workshops and the additional four HGV drivers and road safety officers described in Section 4.4.1. The stakeholders grouped the factors into four categories: (1) driver personal traits; (2) work life and external pressures; (3) in-vehicle technologies; and (4) environmental conditions. Those eliminated during the workshop are represented using strikethrough texts; and additional factors arising from the workshops with stakeholders are in bold text. It is important to note that the factors presented in Figure 4.7 are not ultimate, as other stakeholders who did not take part in our workshops may have different opinions. Therefore, the list could always be updated by interviewing more stakeholders in the future.

Driver personal traits/Driving behaviours
- **Happy,** Sad, Angry
- Tired, **Energetic**
- ~~Driver training~~
- Distracted, **Attentive**
- **Calm**, Aggressive
- **Confident, Insecure**

Work life and external pressures
- Time pressure for delivery
- ~~Time of the day~~
- ~~Day of the week~~
- High Traffic, **Low traffic**
- **Start of shift, midshift, end of shift**

Vehicle characteristic & technologies
- ~~Vehicle weight, age~~
- Driver-facing cameras
- Road-facing cameras
- Multi-cameras
- Absence of cameras
- **Telematics**
- **Absence of telematics**

Environmental conditions
- Rainy, fog, sunny, icy, snowy
- **High and low temperatures**
- **Cloudy, windy, sunrise, sunset**
- **Motorway, divided highway, undivided highway**
- **Intersection,** single/multi-lane urban roads
- Rural roads

Figure 4.7: Contextual factors that impact HGV driving risk extracted from the literature and validated by stakeholders in the HGV industry. The strikethrough factors were eliminated by stakeholders while those in bold fonts represent additional factors identified by stakeholders.

4.5.2 Visualising stakeholder responses

Box-plots and line graphs are utilised to visualise the responses from stakeholders. The box-plots show the group distributions of responses while the line graphs show the individual responses. Each line or dot in the line graphs represents the *(minimum, maximum)* response of each stakeholder and the colours represent their different professions. If a stakeholder's response was very certain (i.e., they selected a single value as their rating), a single point is plotted in the line graph. Each line graph has 93 lines and points for all 93 stakeholders. The line graphs are found in Appendix A.2.

The box-plots show the mode rather than the median because we are interested in the rating provided by majority of the stakeholders in each profession. The following acronyms are used to represent the different professions on the box-plots: HD=HGV drivers, FM=Managers, R=Researchers, and RS= Road safety specialists. The following subsections discuss the box-plots produced by the responses for the different factors.

(a) Drivers' personal traits

The box-plots in Figure 4.8 show the impact of drivers' personal traits on HGV driving performance as suggested by the stakeholders. We observe a less negative effect of feeling tired (Figure 4.8d) suggested by researchers compared to other stakeholders. This may be due to new road safety regulations in the UK that require drivers to take frequent rest breaks [335]. Next, we notice that the majority of researchers suggest that being confident (Figure 4.8j) or insecure (Figure 4.8k) has no impact on driving risk (i.e. mode = 5), which contradicts what the majority of stakeholders from the other groups think. The other stakeholders suggest that being confident has a positive impact, while being insecure has a negative impact on HGV driving.

In an online driving assessment system, the aforementioned driver traits could be detected from driver data using machine learning approaches, as shown in Chapter 3. The information about the detected driver traits are later sent to the risk assessment module for analysis. However, more complex driver traits such as confidence and insecurity are still difficult to detect using machine learning methods due to lack of data about these traits.

(b) Work-life factors and external pressures

Figure 4.9 depicts the distributions of responses from stakeholders on the impact of work-life factors and external pressures. We observe variation between the opinions of drivers compared to managers and road safety professionals for start and end of shift (Figure 4.9a,c) with the mode of drivers' responses at 5. The majority of managers and road safety professionals indicate that the start of a shift has a positive influence on HGV driving and the end of a shift has a negative influence. To better understand the cause of this variability, further investigation by interviewing drivers and managers is required in the future, which is out of the scope of this thesis. Furthermore, time pressure for delivery (Figure 4.9e) is considered to have a strong negative impact (mode = 1) by the majority of managers, who sometimes exert pressure on drivers to deliver on time [336]. This observation stresses the need to consider such factors in the assessment of driving, as HGV drivers sometimes face pressure from their companies to deliver goods on time; literature shows that this could lead to road incidents or accidents [336, 337]. Therefore, for a fairer

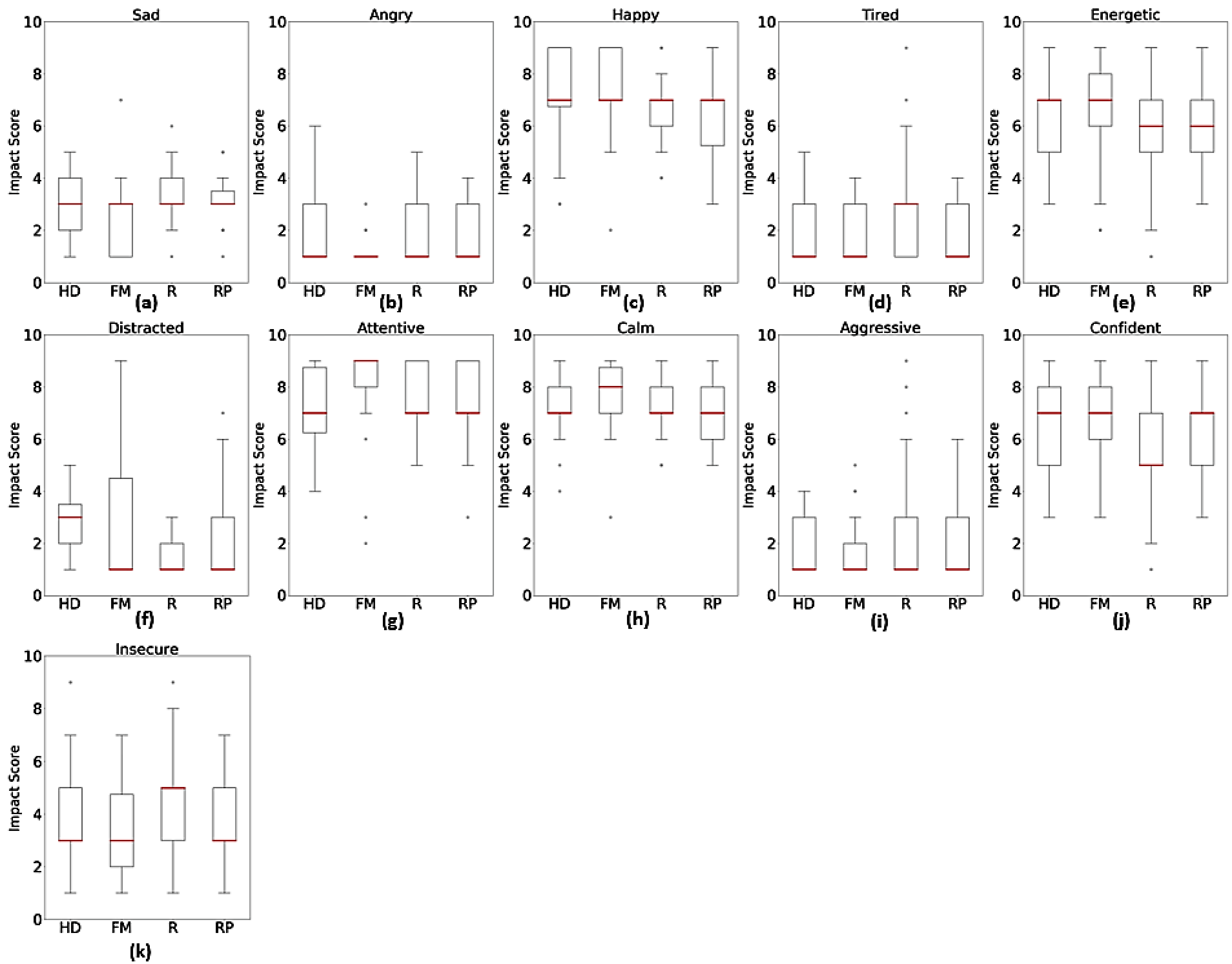


Figure 4.8: Box-plots showing the distribution of responses from drivers, managers, researchers and road safety professionals about the impact of driving behaviours on their HGV driving risk.

assessment of HGV drivers, pressure from their employers cannot be ignored when assessing their driving performance.

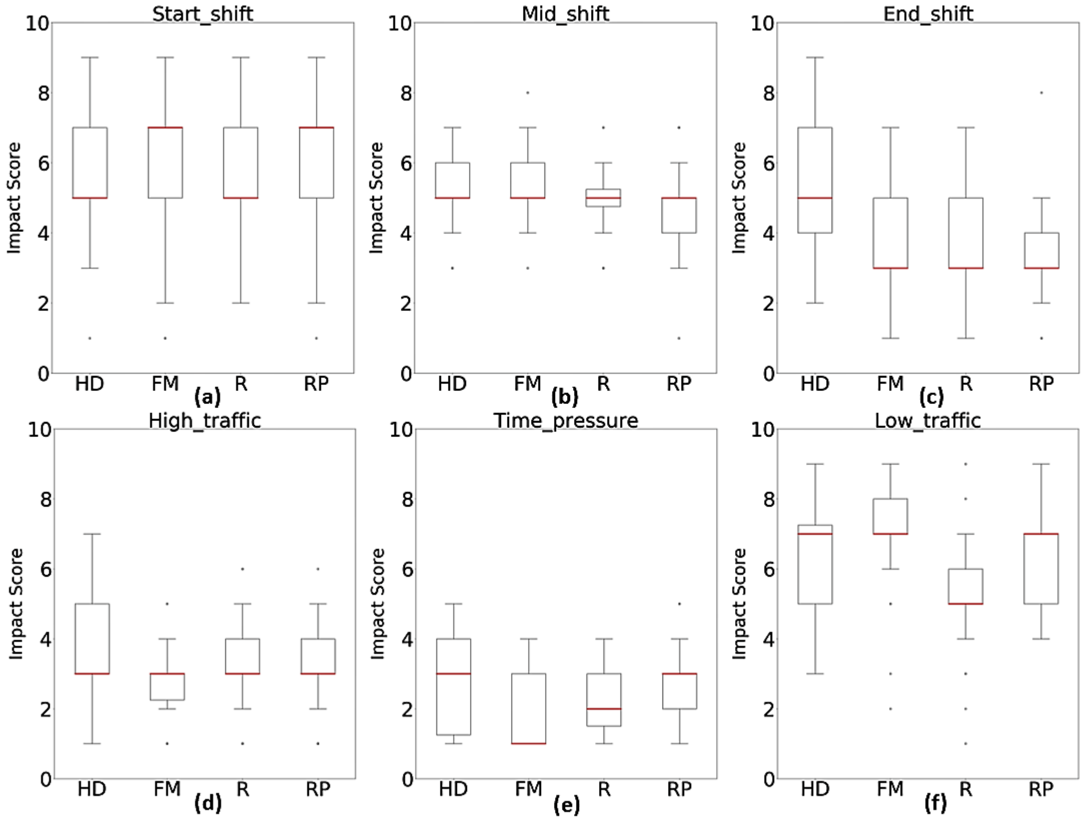


Figure 4.9: Box-plots showing the distribution of responses from drivers, fleet managers, researchers and road safety professionals about the impact of work related factors and external pressures.

In an online driving assessment system, the time of shift (i.e. start, mid and end of shift) can be determined using job dispatch and routing management systems, while traffic state (i.e. high or low) can be automatically recognised from road-facing camera images using computer vision techniques [338, 339] or obtained from location based systems e.g., Google Maps. For time pressure, it is still very difficult to automatically detect.

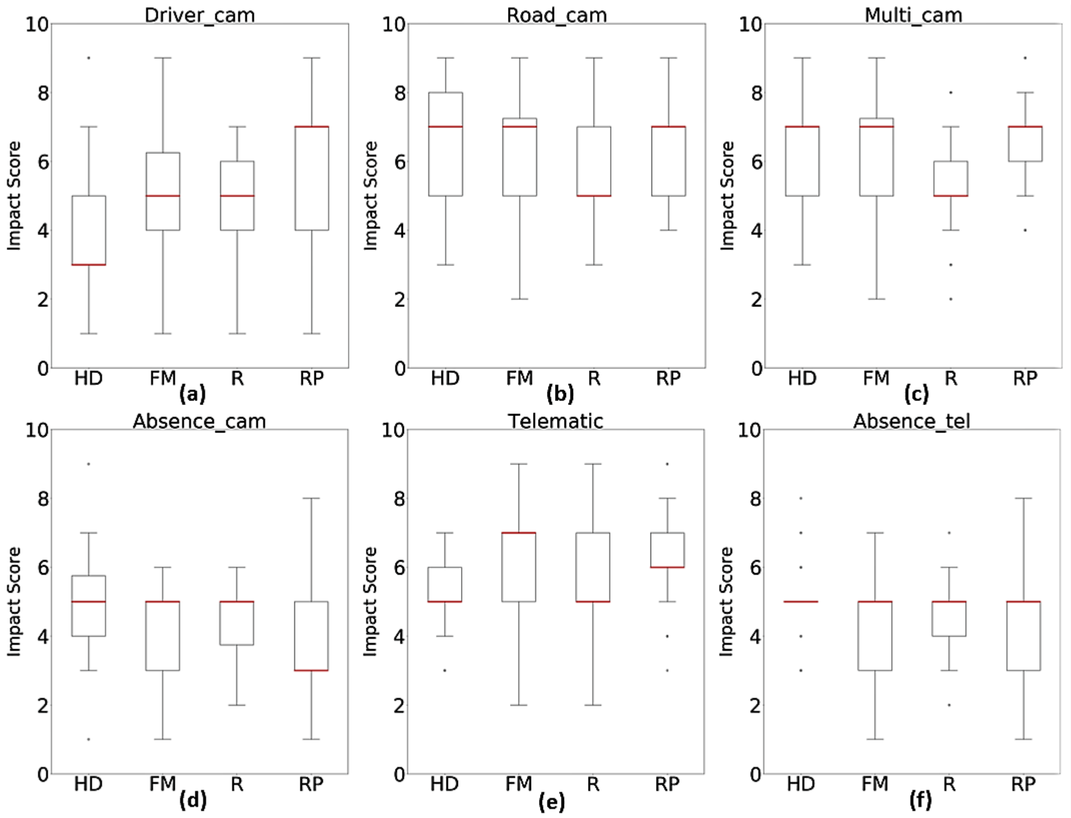


Figure 4.10: Box-plots showing the distribution of responses from drivers, fleet managers, researchers and road safety professionals about the impact of in-vehicle technologies.

(c) In-vehicle technologies

Figure 4.10 depicts the distributions of responses from stakeholders about the impact of in-vehicle technologies. Drivers and road safety professionals disagreed about the impact of driver-facing cameras (Figure 4.10a) as majority of drivers suggest driver-facing cameras have a negative impact on driving, while road safety professionals believe driver-facing cameras have a positive impact. In addition, road safety professionals suggest that the absence of cameras (Figure 4.10d) has a negative impact on driving while majority of other stakeholders suggest no impact. The negative impact of driver-facing cameras suggested by drivers may be a consequence of how the videos or images are being used (e.g. used to penalise drivers) or due to privacy concerns, as we observe positive ratings by drivers for road-facing cameras, which are less intrusive and personal. A system developed with only the opinions of drivers may be inaccurate with regards to the effects of driver-facing cameras. Such discrepancies show the need for a collaborative system to resolve conflicts, where the opinions of different stakeholders are considered.

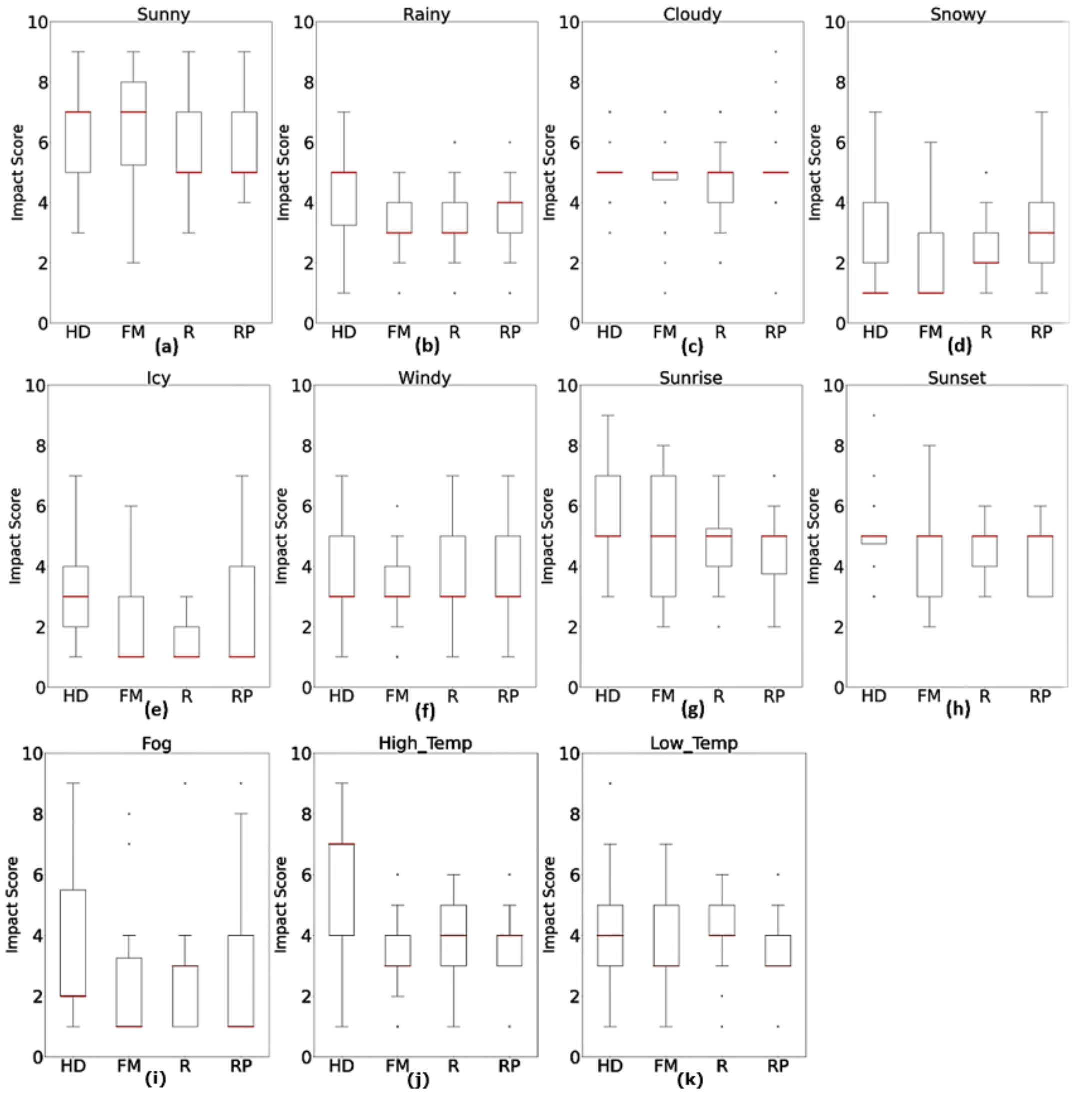


Figure 4.11: Box-plots showing the distribution of responses from drivers, fleet managers, researchers and road safety professionals about the impact of weather conditions.

(d) Environmental factors

Figure 4.11 presents the distributions of responses on the impact of environmental conditions on HGV driving risk. The majority of drivers and fleet managers suggest a sunny weather (Figure 4.11a) has a positive impact (mode = 7) as a sunny weather provides good visibility of the road and other road users, while majority of researchers and road safety professionals suggest it has no impact (mode = 5). Similarly, the majority of drivers indicate a rainy weather (Figure 4.11b) has no impact on driving, while the other stakeholders believe it has a negative impact as visibility and tyre friction are reduced.

Furthermore, we observe negative impact of undivided highways, single-lane urban roads and rural roads suggested by majority of stakeholders (Figure 4.12c,e,g). This is because the roads are single carriageways with no central reservations, mak-

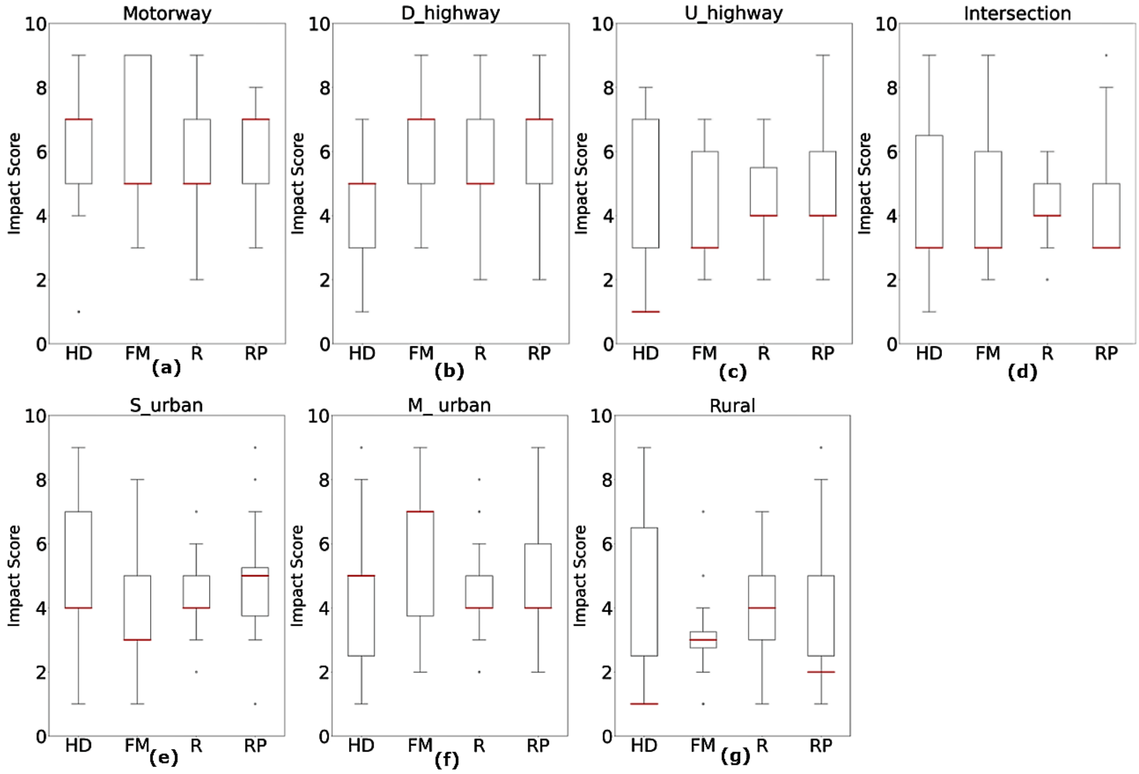


Figure 4.12: Box-plots showing the distribution of responses from drivers, fleet managers, researchers and road safety professionals about the impact of road types.

ing them more susceptible to collisions compared to dual carriageways and motorways [340].

In an online driving assessment system, information about environmental conditions can be automatically obtained using deep learning methods on road-facing images [341, 342, 343], or from online weather data sources such as the Metropolitan Police UK[1] or real-time weather forecast applications.

[1]https://www.met.police.uk/

4.5.3 Agreement amongst stakeholders

Overall, the analysis of the questionnaire data show that stakeholders generally agree on the direction of impact of contextual factors (whether positive or negative impact), but vary in their opinions about the extent of the impact of factors. The variations in magnitude can be attributed to the uncertainties in human perceptions and views as well as differences in roles and expectations, as described in Section 4.2.2. Such variations need to be considered and resolved for the development of a reliable, unbiased and collaborative system.

This section calculates the similarity measures between the responses provided by the different groups of stakeholders regarding the impact of the factors. This is important to show the degree of agreement in views between the different professions. The results are presented in Figure 4.13 with similarity values between 0 and 1, where 0 indicates total disagreement (red) and 1 indicates total agreement (green). We observe from the figure that drivers, researchers and road safety professionals have similarity values above 0.5 for several factors. This is promising as policy enforcers and vehicle operators seem to agree about the effects of most critical factors. However, there are few factors where drivers and researchers have low levels of agreement (less than 40% agreement) i.e., 'distraction', 'driver-facing camera', 'absence of camera' and 'icy weather'. Drivers and road safety professionals have low levels of agreement in 'tired', 'end of shift', 'sunny' and 'sunrise', but agree (above 60% agreement) about 'energetic', 'attentive', 'calm', 'snowy', 'icy' and 'time pressure' with time pressure having the highest agreement (0.833).

Managers show the highest degree of disagreement with other stakeholders, which is precarious as they supervise the drivers and ensure they comply with safety regulations. We observe very high levels of disagreements (below 20%) between managers and drivers for 'angry', 'energetic', 'attentive', 'high traffic', 'time pressure', 'telematics' and 'rainy'. We observe high levels of disagreements between managers and road safety officers (below 20%) for 'attentive', 'time pressure', and 'low traffic'. While managers disagree the most with researchers i.e., 'angry', 'tired', 'attentive', 'aggressive', 'time pressure', 'low traffic', 'absence of camera', 'telematics', 'absence of telematics', 'rainy', and 'cloudy'. This may be due to researchers being mainly exposed to qualitative insights from drivers, thereby, making their insights biased towards drivers.

The disagreements show the need for dialogue across the different professions to attain appropriate, precise and similar perceptions about the influence of the factors. This is crucial to improve communication between the different professions especially managers and drivers, improve company policies, road safety policies, and facilitate the adoption of driver monitoring and feedback technologies.

4.5.4 Perceived impact of contextual factors

Stage 3 of SIA provides a computational approach to aggregate and model the responses from the different stakeholders using the IAA. The advantage of using the IAA approach is : 1) its generalisability to any set-valued dataset, 2) it employs a least commitment strategy; no information is added or removed, and assumptions are kept to a minimum such as the type of fuzzy membership function used, and 3)

Expert group	sad	angry	happy	tired	energetic	distracted	attentive	calm	aggressive	confident	insecure	AVG
Drivers -- Managers	0.355	0.152	0.439	0.267	0.184	0.311	0.132	0.356	0.233	0.483	0.280	0.29 ± 0.11
Drivers -- Researchers	0.647	0.794	0.557	0.563	0.562	0.364	0.739	0.641	0.446	0.412	0.587	**0.57 ± 0.13**
Drivers -- Road safety	0.400	0.598	0.436	0.348	0.637	0.454	0.630	0.600	0.532	0.425	0.512	0.51 ± 0.10
Managers -- Researchers	0.235	0.132	0.416	0.168	0.218	0.262	0.146	0.314	0.122	0.262	0.253	0.23 ± 0.08
Managers -- Road safety	0.211	0.242	0.292	0.707	0.301	0.515	0.187	0.347	0.404	0.294	0.422	0.36 ± 0.14
Researchers -- Road safety	0.532	0.538	0.665	0.230	0.508	0.441	0.715	0.411	0.406	0.559	0.574	0.51 ± 0.13
AVG	0.40 ± 0.15	0.41 ± 0.25	**0.47 ± 0.12**	0.38 ± 0.19	0.40 ± 0.17	0.39 ± 0.09	0.42 ± 0.27	0.44 ± 0.13	0.36 ± 0.14	0.41 ± 0.1	0.44 ± 0.13	

Expert group	start of shift	midshift	end of shift	high traffic	time pressure	low traffic	AVG
Drivers -- Managers	0.356	0.590	0.309	0.199	0.180	0.225	0.31 ± 0.14
Drivers -- Researchers	0.455	0.617	0.531	0.479	0.441	0.559	0.51 ± 0.06
Drivers -- Road safety	0.495	0.556	0.363	0.457	0.833	0.409	**0.52 ± 0.15**
Managers -- Researchers	0.390	0.456	0.273	0.326	0.198	0.133	0.30 ± 0.11
Managers -- Road safety	0.472	0.607	0.386	0.415	0.186	0.174	0.37 ± 0.15
Researchers -- Road safety	0.472	0.579	0.409	0.786	0.508	0.293	0.51 ± 0.15
AVG	0.44 ± 0.05	**0.57 ± 0.05**	0.38 ± 0.08	0.44 ± 0.18	0.39 ± 0.24	0.30 ± 0.15	

Expert group	sunny	rainy	cloudy	snowy	icy	windy	sunrise	sunset	AVG
Drivers -- Managers	0.769	0.151	0.364	0.220	0.299	0.233	0.575	0.452	0.38 ± 0.19
Drivers -- Researchers	0.450	0.491	0.438	0.548	0.366	0.496	0.463	0.430	0.46 ± 0.05
Drivers -- Road safety	0.377	0.590	0.537	0.635	0.698	0.422	0.333	0.425	**0.50 ± 0.12**
Managers -- Researchers	0.429	0.167	0.185	0.273	0.510	0.405	0.452	0.442	0.36 ± 0.12
Managers -- Road safety	0.424	0.210	0.285	0.263	0.384	0.490	0.484	0.540	0.38 ± 0.11
Researchers -- Road safety	0.397	0.636	0.449	0.443	0.408	0.529	0.578	0.501	0.49 ± 0.08
AVG	**0.47 ± 0.13**	0.37 ± 0.2	0.38 ± 0.12	0.40 ± 0.16	0.44 ± 0.13	0.43 ± 0.10	0.42 ± 0.27	0.44 ± 0.13	

Expert group	Fog	High_Temp	Low_Temp	Motorway	D_highway	U_highway	Intersection	S_urban	M_urban	Rural	AVG
Drivers - Managers	0.194	0.238	0.125	0.170	0.348	0.333	0.353	0.208	0.194	0.429	0.26 ± 0.09
Drivers - Researchers	0.117	0.099	0.351	0.224	0.071	0.152	0.109	0.139	0.142	0.052	0.15 ± 0.08
Drivers - Road safety prof	0.272	0.098	0.642	0.185	0.088	0.168	0.461	0.364	0.152	0.097	0.25 ± 0.17
Managers - Researchers	0.069	0.223	0.066	0.073	0.082	0.109	0.078	0.175	0.081	0.073	0.10 ± 0.05
Managers - Road safety prof	0.150	0.320	0.217	0.062	0.131	0.110	0.219	0.281	0.093	0.088	0.17 ± 0.08
Researchers - Road safety prof	0.227	0.365	0.326	0.631	0.562	0.463	0.300	0.373	0.648	0.439	**0.43 ± 0.13**
AVG	0.17 ± 0.06	0.22 ± 0.10	**0.29 ± 0.19**	0.22 ± 0.19	0.21 ± 0.18	0.22 ± 0.13	0.25 ± 0.13	0.26 ± 0.0[illegible]	0.22 ± 0.19	0.20 ± 0.17	

Expert group	driver camera	road camera	multi-camera	absence camera	telematics	absence telematics	AVG
Drivers -- Managers	0.494	0.521	0.357	0.220	0.162	0.256	0.33 ± 0.14
Drivers -- Researchers	0.375	0.614	0.492	0.341	0.460	0.494	0.46 ± 0.09
Drivers -- Road safety	0.474	0.469	0.444	0.426	0.587	0.520	**0.49 ± 0.05**
Managers -- Researchers	0.310	0.399	0.312	0.198	0.135	0.172	0.25 ± 0.09
Managers -- Road safety	0.471	0.424	0.448	0.310	0.221	0.379	0.38 ± 0.09
Researchers -- Road safety	0.411	0.431	0.256	0.483	0.408	0.402	0.40 ± 0.07
AVG	0.42 ± 0.06	**0.48 ± 0.07**	0.38 ± 0.08	0.33 ± 0.10	0.33 ± 0.17	0.37 ± 0.12	

Figure 4.13: Agreement amongst stakeholders opinions for the impact of contextual factors

it provides interpretable representations of human insights and uncertainty levels. First, IAA FSs are developed from the set-valued discrete responses by converting the two discrete points into continuous interval-valued responses with discrete endpoints as described in Section 4.3.3. The IAA FSs model and provide a clear visualisation of the agreement among stakeholders within the same group. For example, the plots on the left in Figure 4.14 (the red plots) show IAA FSs of 'feeling sad' generated from the responses of drivers, managers, researchers and road safety professionals. Examining the IAA FSs, we observe that opinions of stakeholders in all four groups have some level of variability about the impact of feeling sad on HGV driving risk, e.g. drivers' IAA FS ranges from 1 to 5 with the highest agreement at 3 (i.e. $\mu(x) = 1$), while managers' IAA FS ranges from 1 to 4 with an outlier at 7. Also, the FS of managers is more skewed compared to the FS of drivers, meaning managers are more certain that 'feeling sad' has a strong negative impact on HGV driving performance compared to drivers.

The IAA FSs of the different groups are aggregated to obtain the final perceived impact of the factors represented as General Type-2 (zGT2) FSs, as shown by the plot on the right in Figure 4.14 (the green plot). The proposed approach assumes equal weights for all stakeholders. The weights can be adapted in the future, however, providing weights is out of the scope of our research. The resulting zGT2 FSs for all the factors are presented in Appendix A.3. The zGT2 FSs represent the agreement (the colour density of the FS) and variability (the width of the FS) of the opinions of all stakeholders across all groups. The shaded green regions in the zGT2 FSs represent the areas where the IAA FSs overlap, effectively weighting areas with high agreement among stakeholder groups. Darker areas represent higher agreement among stakeholder groups. Thus, as shown in Figure 4.14, the outlier at '7' is given a low weight denoted by the bright shade. The zGT2 FSs can be defuzzified and incorporated into online decision-support systems as shown in Section 4.5.5.

Furthermore, the perceived impact of the contextual factors will better inform HGV drivers about the effects of their personal traits and external conditions on road safety. It will also assist road safety professionals in developing adequate traffic laws and road safety policies that take into consideration the impact of contextual factors, and assist researchers to prioritise the development of intelligent approaches for analysing and characterising negative impact factors. Lastly, HGV transport managers will benefit from the collaborative insights by considering the influence of inevitable external factors (e.g., weather and traffic conditions) and the negative impact of time pressure in their management of drivers.

4.5.5 Fusion and integration of contextual information into online driving assessment systems

This section presents SIA's results of incorporating information about the impact of contextual factors into the assessment of HGV driving performance and risk. The results provided in this section do not demonstrate the accuracy of the framework due to the lack of driving risk data. Rather it shows the importance, reliability and fairness of considering context in the assessment of HGV driving performance and risk. Therefore, the results are presented to motivate debate in the HGV and road

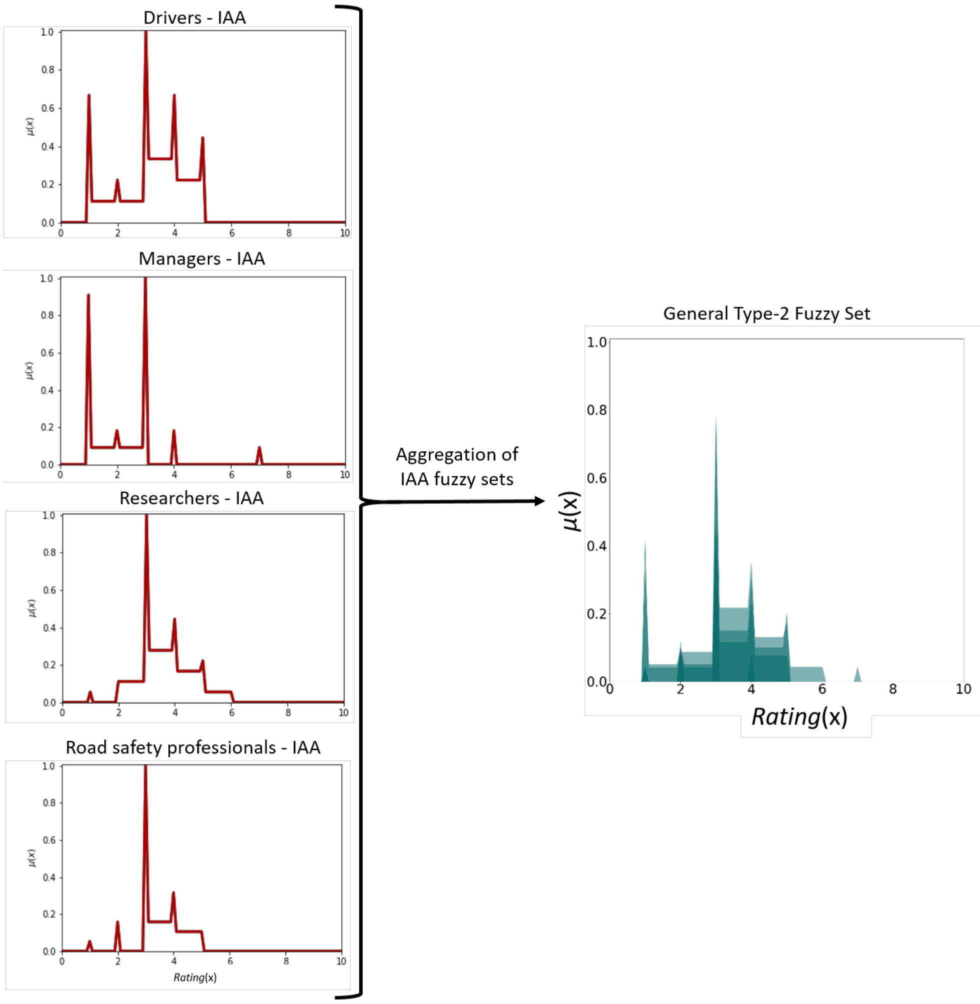


Figure 4.14: The aggregation of IAA fuzzy sets for the impact of 'feeling sad' on HGV driving risk generated from the responses of drivers, fleet managers, researchers and road safety professionals.

safety research communities as well as improve the proposed solution. Stage 4 of the framework is applied to user stories about HGV driving scenarios; designed and validated with help of the stakeholders recruited in Stage 1 of the framework.

(a) Assessing driving performance

Imagine two HGV drivers i.e. Bob and Alice, having exactly the same number of driving incidents after completing journeys of similar duration. Their driving incidents are assessed using CIDA approaches that do not consider context, such as, [14, 305]. Let's assume the approaches rate the drivers' performance as 70%, where 0% represents 'very calm' driving, 50% represents 'moderate' driving, and 100% represents 'very aggressive or reckless' driving. Bob and Alice's manager is not pleased with their assessments and decides to investigate further their driving conditions/context. Bob was driving under little or no time pressure, while Alice was under 'high' time pressure from their company. After incorporating the contextual information, Alice's driving performance was moderated to 48% as shown in Fig. 4.15 and the remaining 22% attributed to the pressure for delivery, while Bob's driving performance was considered riskier, as a better driving performance without the influence of time pressure is expected.

The manager is still not pleased with Bob's assessment and decides to consider the weather conditions during their journeys. Alice was driving in a rainy weather, while Bob was driving in a dry and sunny weather. After incorporating more contextual information into their assessments i.e., weather conditions, Alice driving was further moderated to 40% and the remaining 8% attributed to the inevitable harsh weather conditions, while Bob's driving performance decreased further to 82% as better performance in nice weather conditions is expected. The manager could not think of any more conditions that could have affected their driving performance, and is now able to understand the reasons for the assessments attributed to the drivers in those circumstances. The manager recommends training to Bob and educates them about the consequences of their driving performance to road safety and the associated costs.

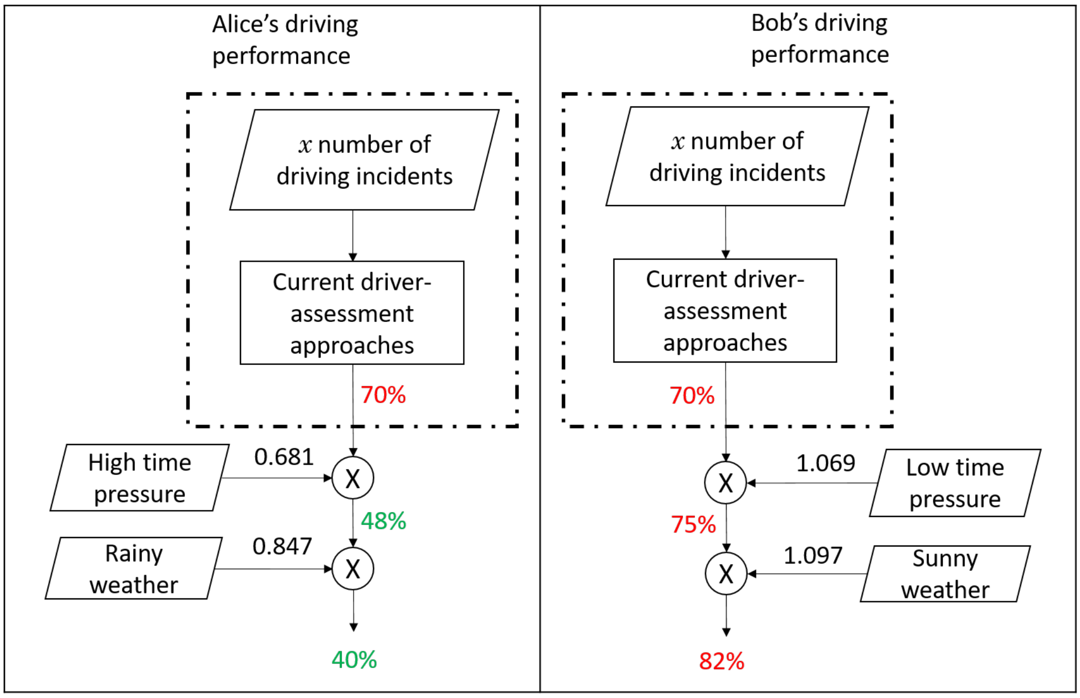


Figure 4.15: A diagram illustrating two examples of incorporating context into the assessment of HGV drivers. 0% represents 'very calm' driving, 50% represents 'moderate' driving, and 100% represents 'very aggressive' driving.

Fig. 4.15 shows the driving performance assessment of Bob and Alice with and without contextual information i.e., influence of time pressure and weather conditions. It can be clearly observe that including the impact of inevitable contextual factors in the analysis of driving performance produces fairer and more useful assessments, which can enable context-specific plans of action to improve safe driving.

Next, a step-by-step process is provided to demonstrate how the FSs generated by SIA are fused and embedded into online driving assessment systems:

(i) Defuzzification of zGT2 FSs: The zGT2 FSs are defuzzified into crisp values for easy fusion with the decisions from CIDA approaches. The defuzzified values of the HGV contextual factors identified by SIA are found in Fig. 4.16. This approach assumes the outputs from CIDA approaches for driver assessment are from 0 to 100, where 0 represents 'very calm' driving and 100 represents 'very aggressive or reckless' driving. Defuzzification compresses the collaborative contextual information into crisp values for fusion with the crisp outputs from the CIDA approaches (Stage 4 of SIA).

CTR defuzzification [330] are employed to reduce the GT2 FSs to crisp information: high time pressure=2.45, low time pressure=7.55, rainy=3.78, and sunny=5.76, where 1 represents 'strong negative impact', 5 represents 'no impact' and 9 represents 'strong positive impact'.

(ii) Fusion and incorporation of defuzzified contextual information: For simplicity, the product aggregation strategy is employed to fuse the defuzzified contextual information, as follows:

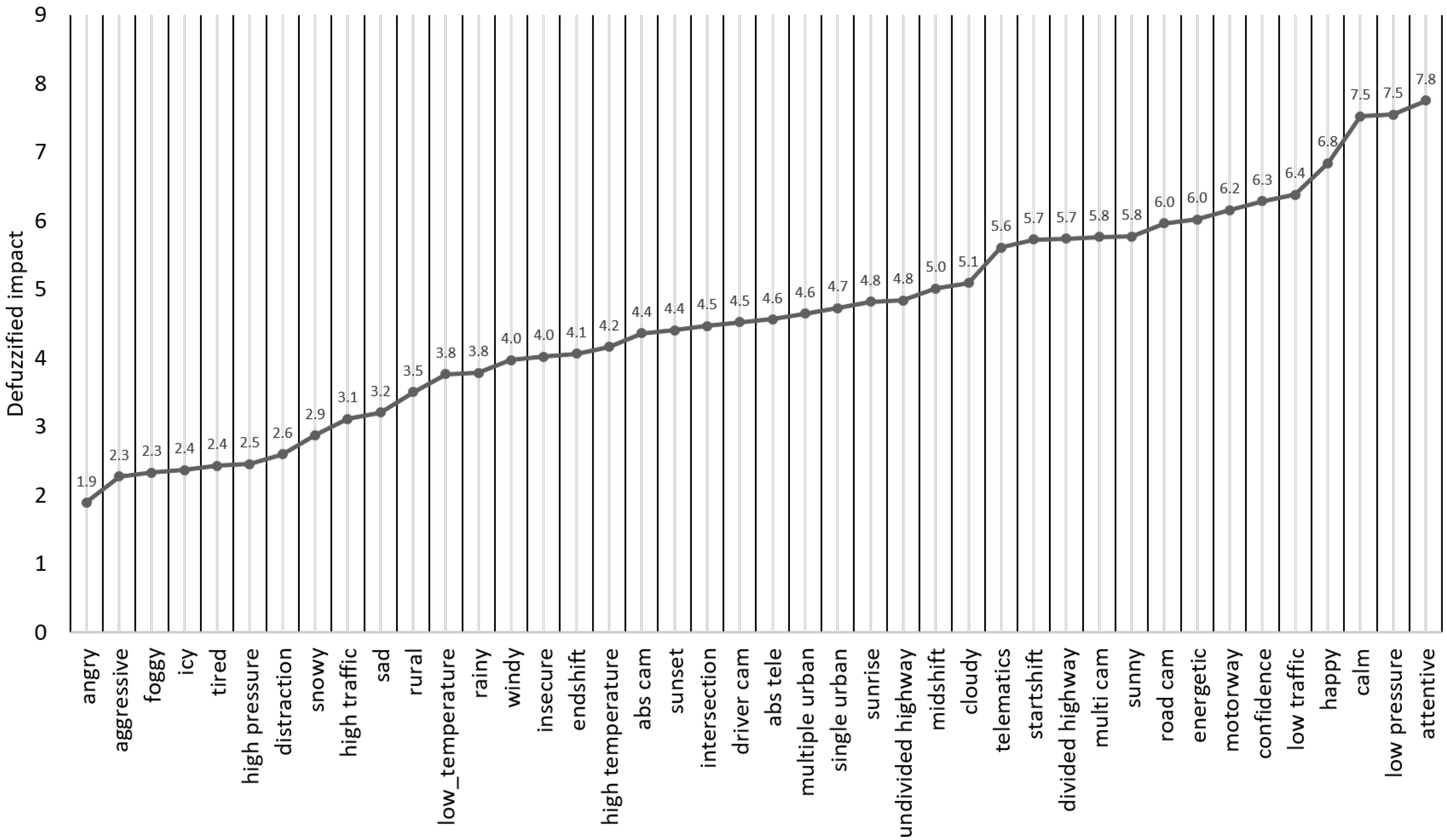


Figure 4.16: Centroid defuzzified values representing the impact of HGV contextual factors on driving performance.

$$CIDA\ driving\ assessment * \prod norm(defuzzified(factors)) \qquad (4.3)$$

First, *[0.5, 1.5]* normalisation is applied to transform the defuzzified values, as follows:

$$x^n = (b-a)\frac{x - xmin}{xmax - xmin} + a \qquad (4.4)$$

Where x^n in $[a,b]$ is the normalised value of x, $xmin$ is the minimum possible value of x (in our case $xmin$=1) and $xmax$ is the maximum possible value of x (in our case $xmax$=9).

[*0.5, 1.5*] is a sensible range for employing the product operator as driving risk is increased or reduced by half for strong negative or positive impact, respectively. The product operator is limited, as it could lead to values greater than 100. The [*0.5, 1.5*] normalisation of the defuzzified values are as follows:

- norm(high pressure=2.45)$_{[0.5,1.5]}$ = 0.681
- norm(low pressure=5.55)$_{[0.5,1.5]}$ = 1.069
- norm(rainy=3.78)$_{[0.5,1.5]}$ = 0.847
- norm(sunny=5.76)$_{[0.5,1.5]}$ = 1.095

The resulting normalised values are fused with the outputs from the intelligent system as follows:

- Bob's driving performance = 70%*1.069*1.095=81.92%
- Alice's driving performance = 70%*0.681*0.847=40.37%

The above examples demonstrate the application of SIA in moderating the decisions of CIDA approaches to produce fairer and more meaningful driving performance assessments.

(b) Driving risk assessments

Stage 4 of SIA can also be applied to provide online driving risk assessments that consider the impact of detected driver traits and perceived external conditions on road safety, which is the main aim of this research. For example, imagine five HGV drivers driving with the following driving styles in different external conditions:

- Tom: *aggressive; rainy; high traffic*
- Jerry: *calm; rainy; high traffic*
- Alice: *aggressive; rainy; low traffic*
- Bob: *calm; rainy; low traffic*

- Jim: *calm; sunny; low traffic*

The [*0.5, 1.5*] normalised defuzzified impact values for the factors are:

- norm(aggressive=2.27)$_{[0.5,1.5]}$ = 0.659
- norm(calm=7.52)$_{[0.5,1.5]}$ = 1.315
- norm(high traffic=3.11)$_{[0.5,1.5]}$ = 0.764
- norm(low traffic=6.38)$_{[0.5,1.5]}$ = 1.172
- norm(rainy=3.78)$_{[0.5,1.5]}$ = 0.847
- norm(sunny=5.76)$_{[0.5,1.5]}$ = 1.095

Journeys are assumed to start with a risk level of 50% (where 0% represents 'very low' risk, 50% represents 'moderate' risk, and 100% represents 'very high' risk) and depending on the impact of perceived factors, the risk level can increase or reduce during the journey. Using the normalised defuzzified values, the risk of the above driving scenarios can be computed using the following equation:

$$Latest\ risk\ level * \frac{1}{\prod norm(defuzzified(factors))} \tag{4.5}$$

CIDA approaches (e.g., [15]) that only consider drivers' driving style in their assessment of driving risk, produce the following assessments:

- Tom's driving risk = $50\% * \frac{1}{0.659} = 75.9\%$
- Jerry's driving risk = $50\% * \frac{1}{1.315} = 38.0\%$
- Alice's driving risk = $50\% * \frac{1}{0.659} = 75.9\%$
- Bob's driving risk = $50\% * \frac{1}{1.315} = 38.0\%$
- Jim's driving risk = $50\% * \frac{1}{1.315} = 38.0\%$

By including information about the impact of external factors on road safety i.e., weather and traffic conditions, the following assessments are produced:

- Tom's driving risk = $50\% * \frac{1}{0.659*0.847*0.764} > 100\%$
- Jerry's driving risk = $50\% * \frac{1}{1.315*0.847*0.764} = 58.8\%$
- Alice's driving risk = $50\% * \frac{1}{0.659*0.847*1.172} = 76.4\%$
- Bob's driving risk = $50\% * \frac{1}{1.315*0.847*1.172} = 38.3\%$
- Jim's driving risk = $50\% * \frac{1}{1.315*1.095*1.172} = 29.6\%$

The results show that without considering the driving circumstances, drivers with similar driving styles produce the same level of driving risk. For example, Bob, Jim and Jerry produce the same driving risk even though it is certain that driving in rainy weather or high traffic conditions is more risky than driving in sunny weather or low traffic conditions. Obviously, it is appropriate that Bob and Jerry keep a calm driving style in the negative external conditions, however, Bob and Jerry's driving circumstances are clearly different from Jim's and should be considered when assessing their driving risk. When the impact of external factors are considered in the assessment of driving risk, we observe that Jerry's driving scenario (calm; rainy; high traffic) is considered more risky than Bob's (calm; rainy; low traffic) and Bob's circumstance is more risky than Jim's (calm; sunny; low traffic). Similarly, Tom's driving circumstance (aggressive; rainy; high traffic) is considered more risky than Alice's (aggressive; rainy; low traffic), as we will expect.

4.5.6 Limitations of SIA

This section provides some limitations and challenges of SIA that require improvement and further analysis.

- The framework does not capture the synergy and interactions between drivers' personal traits and external factors during driving, which could potentially lead to highly correlated factors e.g., a driver's driving style may be affected by their affective state. As revealed by the psychological theories on driver behaviour in Chapter 2, contextual factors do not occur independently during driving. They occur simultaneously and interact with each other to impact driving risk. Therefore, in order to produce more accurate and realistic assessments of driving risk, these interactions or relationships need to be effectively captured and modelled.

- Even though the framework captures uncertainties in stakeholders' perception about the impact of contextual factors, it fails to incorporate the uncertainties into the assessment of driving risk. This is due to its defuzzification process that reduces the variability in expert opinions (represented as zGT2 FSs) into crisp values. To produce more robust and trustworthy driving risk assessments that consider variability in human knowledge and opinions, the zGT2 FSs need to be incorporated into driving risk assessments without loss of the uncertainty information.

- The framework does not consider the uncertainties in the assessment of driving risk, such as, imprecision in the definitions of driver traits, uncertainty in the decisions of AI models, and variability in the subjective evaluation of driving risk by multiple experts. Therefore, for a more trustworthy and interpretable assessment of driving risk, the aforementioned uncertainties in information need to be effectively captured, modelled and presented.

- Lastly, several assumptions are made by the framework that require further analysis and validation, such as the [0.5, 1.5] normalisation domain and the 50% journey start risk level.

4.6 Summary

This chapter introduced a new framework called SIA that engages with stakeholders in the HGV sector to effectively capture and understand the impact of HGV contextual factors on road safety. This framework is introduced due to the absence of contextual information in current driver data sources and CIDA approaches. The framework, therefore, complements current empirical data by incorporating information about the impact of contextual factors into the assessment of HGV driving risk. The goal is to produce more reliable, fair, meaningful and comprehensive assessments of HGV driving performance and risks based on the impact of drivers' personal traits and external conditions. SIA consists of four main stages: 1) identification of relevant contextual factors that affect HGV driving by stakeholders; 2) design and distribution of questionnaires to key stakeholders in the sector to capture insights about the influence of contextual factors; 3) analysis and modelling of contextual information obtained from the stakeholders; and 4) fusion and incorporation of contextual information into the assessment of driving risk. Based on the analysis of the responses captured from 93 HGV stakeholders (i.e. 20 HGV drivers, 23 researchers, 24 HG managers, and 26 road safety professionals), we noticed that stakeholders agree on the direction of impact of factors (whether positive or negative impact), but vary in their opinions about the extent of the impact. The differences in opinions can be attributed to variability in human views due to different roles, knowledge, experiences and goals. SIA effectively captures, models and represents the variability among stakeholders' views using FSs based on IAA.

Furthermore, the framework measured the similarity between the opinions of the different professions and identified the highest degree of disagreement between managers and other professions. The disagreements among professions show the need for dialogue across the different professions to attain appropriate, precise and similar perceptions of the influence of the factors. This is crucial to improve communication between the different professions especially managers and drivers, improve company policies, road safety policies, and facilitate the adoption of driver monitoring and feedback technologies. The contextual information obtained from the 93 stakeholders were modelled and aggregated using IAA and zGT2 FSs. The last stage of SIA consists of an information fusion approach to fuse the contextual information and moderate decisions from CIDA approaches. Application of the approach using user stories about realistic HGV driving scenarios, designed with the help of stakeholders, shows the ability of SIA to embed stakeholders' inputs and provide a fairer and more meaningful assessments of driving performance and risks that takes into consideration external circumstances.

Although this framework has brought many benefits in understanding the impact of contextual factors and incorporating context into driving risk assessment, it is limited in the following aspects: (1) it does not capture the synergistic interaction between drivers' personal traits and external factors during driving; (2) it does not consider variability of expert inputs in the assessment of driving risk; (3) it does not consider imprecision in the definition of contextual factors; (4) it does not consider uncertainty in decisions from AI models; and (5) it has multiple assumptions.

The next chapter addresses these limitations by extending SIA to capture, model and embed information about the synergy and interactions between drivers' personal traits and external conditions. The extended framework also models the uncertainty and interpretation of driving risk assessment using a hierarchical rule-based fuzzy inference system.

Chapter 5

Online Context-Aware Driving Risk Assessment

5.1 Introduction

Chapter 3 introduced an intelligent multi-modal driver characterisation framework to predict the main facets of driver behaviour that impact road safety i.e., driving styles, driving postures and affective states. Chapter 4 introduced a framework called **S**takeholder-supported **I**ntelligent driving **A**ssessment (SIA) that engages with stakeholders to capture and incorporate information about the impact of detected driver traits and perceived external factors on road safety into the assessment of heavy goods vehicle (HGV) driving performance and risk. Although SIA has brought many benefits in understanding the impact of HGV contextual factors and providing collaborative, comprehensive and fairer HGV driving assessments; it is limited in the following aspects: (1) it does not capture the interactions and synergy between contextual factors, which could potentially lead to highly correlated factors, unstable and overestimated assessments; (2) it does not consider ambiguity in the definition of driving behaviours; (3) it does not consider uncertainty in the information about driver traits and external factors produced by artificial intelligence (AI) models; (4) it does not consider variability in experts' subjective views about the interaction of factors and assessment of driving risks; and (5) it makes several assumptions that require further analysis and validation.

This chapter tackles the above limitations and answers the research questions: "how can a reliable driving risk assessment system that considers the real-world characteristics of the driving environment be developed, taking into consideration the lack of comprehensive driving risk datasets?" and "how can the reliability and effectiveness of the driving risk assessment system be evaluated taking into consideration the lack of multi-modal data?" The chapter proposes an extension of SIA called **S**takeholder-supported **I**ntelligent **F**uzzy driving **A**ssessment (SIFA) that integrates expert inputs about the interactions and causal relationships of contextual factors using linguistic fuzzy IF-THEN rules, while addressing uncertainty or ambiguity in the factors using fuzzy sets (FSs). To ensure that interactions between contextual factors are easy to manage and understand, the framework employs a hierarchical fuzzy approach [344]. The resulting hierarchical rule-based fuzzy inference system from SIFA receives as inputs; information about detected driver traits from driver characterisation AI models (e.g., contribution of Chapter 3) and assesses their synergistic effects together with perceived external conditions on road safety. Information about external conditions (e.g., weather, road and traffic conditions) can be automatically obtained using AI methods on road-facing camera images [341, 342, 343] or from online data sources such as the Metropolitan Police UK[1].

The remainder of the chapter is organised as follows. Section 5.2 provides an overview of fuzzy expert systems and hierarchical fuzzy inference systems, which are the fundamental elements of SIFA. Section 5.3 introduces the different stages of SIFA. Section 5.4 presents the experimental design of the application of SIFA in assessing HGV driving risk. Results and discussions of the framework's application are presented in Section 5.5 using realistic HGV driving scenarios. Section 5.6 summarises the chapter.

5.2 Background

5.2.1 Overview of fuzzy expert systems

Fuzzy expert systems (FESs) integrate human expert knowledge into intelligent systems while addressing imprecision and uncertainties in information, human language and decisions [18]. FESs also provide a natural and easy way to fuse heterogeneous information e.g., information about driver traits and external factors. The expert knowledge is captured and modelled using IF-THEN rules [19]. The uncertainties are captured and modelled using linguistic terms with soft boundaries (FSs). FESs consist of three main stages: 1) fuzzification; 2) knowledge base acquisition; and 3) inference engine.

Before describing the different stages of a FES, the definition of a FS is revisited as described in Section 2.3.4 (page 21). That is, a fuzzy set is a function that assigns inputs to membership degrees between 0 and 1.

In fuzzification, an input (e.g. a signal from a sensor or information from an intelligent system) is converted into singleton or non-singleton FSs. A singleton FS has only one element with a membership value of 1, while a non-singleton FS has a set of elements with their degrees of membership between 0 and 1. Figure 5.1 shows three non-singleton FSs represented as 'low', 'moderate' and 'high' (blue, orange and green plots, respectively), and one singleton FS represented by the black vertical line with its only member as '35' having a degree of 1. In the case where non-singleton FSs are defined to capture uncertainties, fuzzification is the intersection between the singleton FS produced by an input value and the non-singleton fuzzy sets. For example, using the non-singleton FSs defined in Figure 5.1, an input of '35' generates the singleton fuzzy set (represented by the black vertical line) and intersects with the non-singleton fuzzy sets to produce its membership in the non-singleton fuzzy sets i.e. 0 degree of membership in 'low', 0.8 in 'moderate' and 0.2 in 'high'. In a fuzzy logic system (FLS), the non-singleton FSs for the inputs and outputs are defined by domain experts or from historical data.

[1]https://www.met.police.uk/

In the rule knowledge base acquisition stage of FESs, domain experts define the IF-THEN rules. When multiple experts are involved in developing the set of rules, the variability in their views need to be effectively captured and aggregated [19]. More details on how the variability in expert inputs are captured and handled during rule specification is provided in Section 5.3.4.

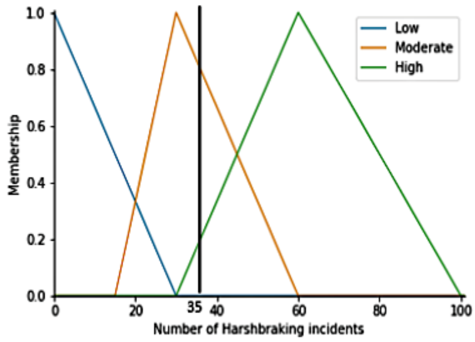


Figure 5.1: Illustration example of the intersection of singleton (black vertical line) and non-singleton fuzzy sets (low, moderate and high plots).

The inference engine aggregates the fuzzified input FSs obtained from fuzzification stage using fuzzy operators in the set of fuzzy rules obtained from the rule knowledge base acquisition stage to produce some expert conclusion or decision. A simple example is used to illustrate the inference process in fuzzy systems. Imagine a driving risk assessment system with only two features: (1) number of harsh braking (HB) incidents; and (2) weather condition, and the following rules that capture the interaction between the features and their impact on driving risk:

***Rule 1**: **IF** HB is "low" **AND** weather is "good" **THEN** Driving risk is "low"*
***Rule 2**: **IF** HB is "low" **AND** weather is "poor" **THEN** Driving risk is "moderate"*
***Rule 3**: **IF** HB is "moderate" **OR** 'high' **AND** weather is "good" **OR** "poor" **THEN** Driving risk is "high"*

Using the FSs and membership function (MF) of HB in Figure 5.1, the above rules generate rule strengths for the different FSs during inference. The rule strengths and their respective output FSs capture the uncertainty and imprecision of the system's conclusion of driving risk. The rule strengths are computed using the fuzzy operators found in the rules e.g., 'AND' (minimum of membership degrees) and 'OR' (maximum of membership degrees) found in Rules 1, 2 and 3 above.

Imagine the perceived number of HB equals '35' and the weather is 'poor', how can we fuse these heterogeneous information about HB incidents and weather to produce an assessment of driving risk? Let's assume singleton FSs for weather with maximum membership for the different weather conditions, and assume the non-singleton FSs in Figure 5.1 for low, moderate and high HB. That is;

$$\begin{aligned} &\mu_{low}(35) = 0.0 \\ &\mu_{moderate}(35) = 0.8 \\ &\mu_{high}(35) = 0.2 \end{aligned} \tag{5.1}$$

The resulting rule strengths show that the driving risk of the occurrence of '35' harsh braking incidents in a 'poor' weather condition is definitely not 'low' nor 'moderate' but has an 80% likelihood of 'high' risk, as follows:

$$\begin{aligned} &Rule1 : Strength = min(0.0, 0.0) = 0.0, Risk = \text{“}low\text{”} \\ &Rule2 : Strength = min(0.0, 1.0) = 0.0, Risk = \text{“}moderate\text{”} \\ &Rule3 : Strength = min(max(0.8, 0.2), max(0.0, 1.0)) = 0.8, Risk = \text{“}high\text{”} \end{aligned} \tag{5.2}$$

5.2.2 Challenges in the development of fuzzy expert systems

There are two main challenges in the development of FESs that need to be considered when designing a system with many contextual factors and multiple domain experts.

(a) Variability between expert views

As earlier mentioned, the causal relationships and interactions between inputs and outputs are captured from domain experts using IF-THEN rules. The knowledge from the domain experts are likely to differ due to differences in precision, perception, experiences and expectations. Thus, it is expected that different experts — even though they have similar roles — may provide different viewpoints about the relationships between variables. In addition, experts with different roles may have varying viewpoints due to their distinct responsibilities and contrasting interests. Therefore, varying rules are captured from experts, which require an effective approach to be modelled, aggregated and incorporated into the system. In Wagner *et al.* [19], the authors proposed a novel survey-centric methodology which enables the capture, aggregation and incorporation of subjective input from multiple domain experts about the interaction between variables into a fuzzy system. Their approach is adopted in SIFA to deal with variability between expert views.

(b) High number of fuzzy rules

In complex systems with many variables, the relationships between the variables increase exponentially. This leads to a huge set of complex rules (i.e., increased number of input conditions and heterogeneous information) that are difficult to manage, understand, reason and fuse; particularly when we depend on experts to define the relationships and when we depend on the outcomes to make informed decisions. For example, imagine one rule with 8 variable conditions and another with 3 variable conditions. The interactions in the second rule with 3 conditions are easier to reason and combine compared to the first rule having 8 conditions at the same time. Soua *et al.* [345] proposed a supervised learning method to reduce the number of rules and antecedent conditions. However, their approach requires labelled data, which is not available in the driving assessment domain. Hierarchical fuzzy approach [225, 346] is an alternative to resolve this challenge, which does not require labelled data, but rather it formats the structure of the entire fuzzy system. Section 5.2.3 provides an overview of hierarchical fuzzy systems and Section 5.4.1 illustrates our adoption of the approach in SIFA.

5.2.3 Overview of hierarchical fuzzy systems

Standard FLSs are made up of a single set of rules with all possible interactions between the inputs. For example, Figure 5.2 shows a standard FLS with four input variables where all the fuzzified inputs are combined into a single system to produce an output. If the variables have three FSs each, there will be 81 possible combinations of input FSs, i.e., number of rules, and each rule has four antecedent conditions. Generally, the total number of rules of a FLS (R_{FLS}) with n input variables and m FSs for each variable is given by the following equation:

$$R_{FLS} = m^n \tag{5.3}$$

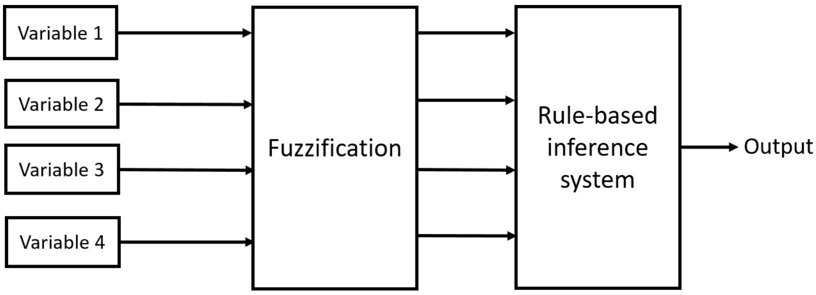


Figure 5.2: A standard fuzzy logic system with four inputs and one output

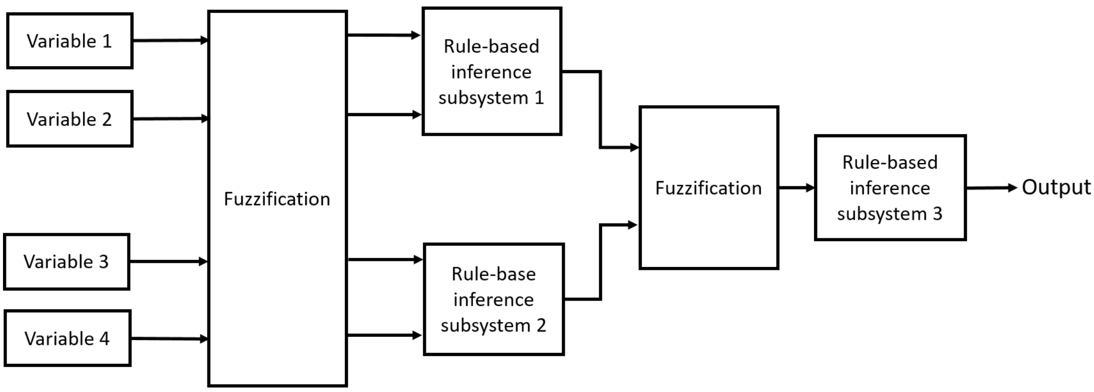


Figure 5.3: A hierarchical fuzzy system made up of three intermediate fuzzy logic subsystems.

Hierarchical Fuzzy Systems (HFSs) group input variables into subsystems representing different facets of the system [225, 346]. The subsystems are lower-dimensional FLSs, which are easier to manage and understand. Figure 5.3 shows a HFS with four input variables and one output. The input variables have been grouped into two subsystems i.e. subsystems 1 and 2, whose outputs interact in subsystem 3 to produce the final output. From a functional standpoint, the FLS and HFS in Figures. 5.2 and 5.3, can be represented as follows:

$$y_{FLS} = F(v_1, v_2, v_3, v_4) \tag{5.4}$$
$$y_{HFS} = F(F(v_1, v_2), F(v_3, v_4)) \tag{5.5}$$

where v_i is the *ith* input variable, and F is a standard FLS.

If we assume three FSs for each variable and three FSs for each intermediate output (i.e. outputs from subsystems 1 and 2), the total number of rules in the HFS is 27 (9 rules from subsystem 1 + 9 rules from subsystem 2 + 9 rules from subsystem 3) and each rule having only two antecedent conditions. The number of rules and antecedent conditions generated by HFSs are therefore significantly lower than those generated by standard FLSs.

The reduction in rules and input conditions by HFSs enable the capture of more reliable and transparent causal relationships as experts have to analyse fewer information and conditions in each subsystem. It also reduces the computational complexity of the system. Furthermore, human knowledge is generally organised hierarchically into levels of abstraction, especially when dealing with complex information [347] e.g., if a human is presented with information about dogs, forks, goats,spoons, cows, and knifes; they will naturally categorise the information into animals and cutlery at the first level of abstraction before further processing. Therefore, when dealing with complex heterogeneous information that describe different concepts of a system, grouping information of similar abstraction makes the system easier for humans to understand, process, evaluate and improve. Section 5.4.1 describes how a hierarchical fuzzy approach is implemented in SIFA to reduce system complexity and increase interpretability.

5.3 Stakeholder-Supported Fuzzy Driving Assessment Framework

This section introduces our new stakeholder-supported driving assessment framework called SIFA. It extends the SIA framework introduced in Chapter 4 by engaging with stakeholders to capture information about the impact of the co-occurrence and interaction of contextual factors on road safety, and employs a HFS to provide decomposable and reliable assessments of HGV driving risk. In addition, SIFA produces a database of driving rules that show the level of road safety risk for driving behaviours in different external conditions.

Figure 5.4 shows a diagram of the five stages of SIFA. The stages are described below:

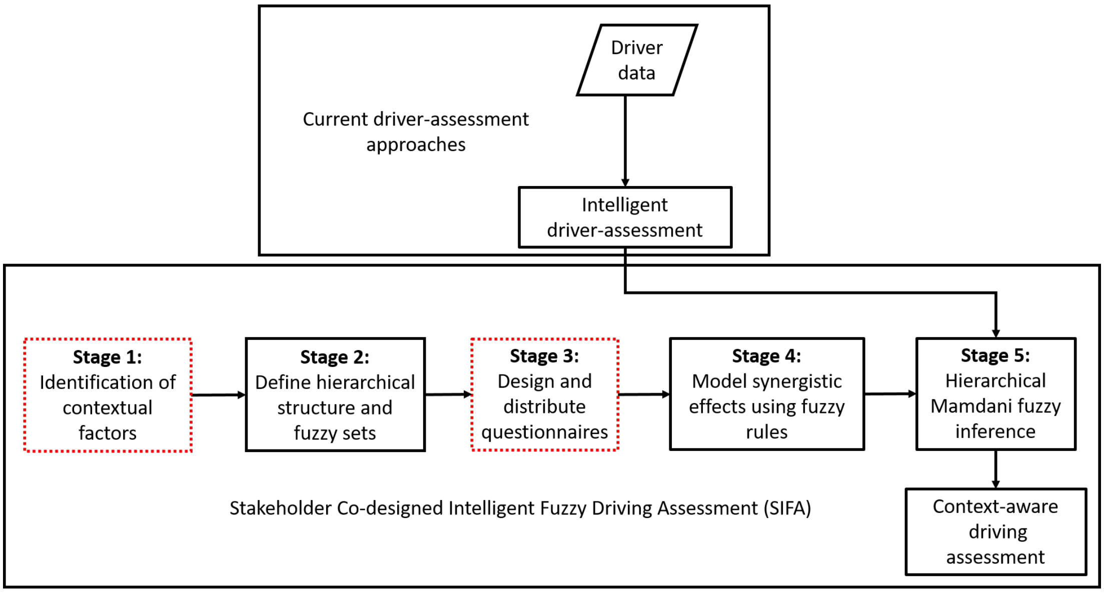


Figure 5.4: A diagram showing the different stages of SIFA framework. Similar stages between SIFA and SIA are indicated with red dashed bounding boxes.

5.3.1 Stage 1: Identification of contextual factors

The first stage of SIFA is similar to Stage 1 of SIA described in Section 4.3.1 (page 77). The stage involves compiling a list of contextual factors that affect driving performance based on the literature. Subsequently, consultative workshops are organised with stakeholders in the industry to validate, update and capture new factors that affect safe driving.

5.3.2 Stage 2: Design hierarchical structure and fuzzy sets

Stage 2 defines the hierarchical structure of the system with the help of domain experts (stakeholders). This is achieved by grouping the factors identified in Stage 1 into similar themes (subsystems) representing different facets of the system. For example, in SIA, the stakeholders grouped HGV contextual factors into four themes representing different aspects of HGV driving: (1) driver personal traits; (2) work life and external pressures; (3) vehicle characteristics and technologies; and (4) environmental conditions. Subsequently, the experts determine how the different subsystems interact to produce the final assessment of driving risk. This defines the thematic subsystems of the hierarchy.

After defining the hierarchical structure of the system, the fuzzy sets (FSs) for the contextual factors and driving risk outputs are defined. The FSs are human-understandable linguistic terms for the different categories or classes of the factors and driving risk decisions. For example, drivers' affective states can be classified into three FSs: 'negative', 'neutral' and 'positive', and driving risk can be classified into 'low', 'moderate' and 'high'. The FSs and their respective membership functions (MFs) model imprecision in the description of factors, smooth transitions between categories of factors, and uncertainty in information produced by AI approaches as shown in Section 5.4.2.

5.3.3 Stage 3: Design and distribute questionnaires

This stage designs questionnaires to capture stakeholders' views about the impact of the interactions between contextual factors on driving risk. These interactions are represented as IF-THEN rules consisting of the linguistic terms (FSs) defined in Stage 2 e.g., a sample interaction could be "IF driver's affective state is 'negative' AND driving style is 'aggressive' THEN driving risk is 'very high' ". The questionnaires are designed to capture imprecision or indecision in the opinions of stakeholders using, for instance, interval-valued response-format questionnaires [232]. Similar to Stage 2 of SIA (Section 4.3.2), the design and wordings of the questions are improved with the help of the stakeholders recruited in Stage 1. The questionnaires are distributed to a wide cohort of key stakeholders in the domain.

5.3.4 Stage 4: Model synergistic effects of contextual factors

This stage models the questionnaire responses about the causal relationships between contextual factors and driving risk. To model and aggregate the variability in stakeholders' views, Wagner *et al.* [19] process of generating weighted fuzzy logic rules from uncertain responses is adopted. The width of a given interval response represents the certainty of the response as described in Section 4.2.4. Each question in the questionnaire produces N interval responses, where N represents the total number of participants that completed the questionnaire. Wagner *et al.* [19] process is described as follows:

1. First, responses of each question (i.e., rule) are aggregated into two summary lists depending on their level of uncertainty i.e., M^{Cert} and M^{Uncert}. M^{Cert} aggregates responses with no uncertainties i.e. stakeholders provided a precise rating for the interaction of contextual factors, while M^{Uncert} aggregates uncertain responses represented as intervals. The sizes of M^{Cert} and M^{Uncert} are equal to the number of options in the questionnaire's rating scale, and each value in the lists represents the occurrence of that particular rating in the response. For example, Figure 5.5 shows summary lists produced from five responses i.e. two certain responses(i.e. (3,3) and (4,4)) and three uncertain responses (i.e. (3,4), (3,5) and (2,3)). The output (or rating scale) has five labels i.e., 1 to 5.

(3, 3) Labels : [1, 2, 3, 4, 5]
(3, 4)
(3, 5) $\longrightarrow$ $\boldsymbol{M^{Cert}}$ **: [0, 0, 1, 1, 0]**
(4, 4)
(2, 3) M^{Uncert} : [0, 1, 3, 1, 1]

Figure 5.5: An example illustrating how summary lists M^{Cert} and M^{Uncert} are generated from certain and uncertain responses. Certain responses are indicated in bold.

2. Subsequently, the summary lists are utilised to compute rule weights, as follows:

$$W_{i,j} = \frac{M_{i,j}^{Cert} \times w^{Cert} + M_{i,j}^{Uncert} \times w^{Uncert}}{w^{Cert} + w^{Uncert}} \tag{5.6}$$

   Where i is number of rules, j is the number of output FSs or labels, $W_{i,j}$ is the rule weight for rule i and label j, $M_{i,j}^{Cert}$ is the value of label j in the certain summary list produced by rule i, and $M_{i,j}^{Uncert}$ is the value of label j in the uncertain summary list produced by rule i. $w^{Cert} \in [0,1]$ is the weight for certain responses, and $w^{Uncert} \in [0,1]$ is the weight for uncertain responses. w^{Cert} and w^{Uncert} reflect how much emphasis is to be given to certain and uncertain responses and is commonly determined by the decision makers.

   To demonstrate the application of the rule weight equation (Equation 5.6), the example in Figure 5.5 is utilised, and the following assumptions are made: $w^{Cert} = 1$ and $w^{Uncert} = 0.5$. The resulting weights produced by the certain and uncertain responses for the five output labels are 0, 0.33, 1.67, 1, and 0.33.

3. Lastly, the rule weights are normalised to (0,1) for integration into fuzzy logic systems as degrees of membership lie between 0 and 1. Therefore, the weights obtained in the previous stage (i.e., 0, 0.33, 1.67, 1, 0.33) are normalised to 0, 0.19, 1, 0.59, 0.19. The normalised weights are used to modify the rule strengths during inference, as follows:

$$f'_{i,j} = f_{i,j} \times W_{i,j} \tag{5.7}$$

   Where $f_{i,j}$ represents the standard rule strength for rule i and label j obtained during inference. $W_{i,j}$ represents the normalised weighted rules for rule i and label j, and $f'_{i,j}$ represents the resulting weighted rule strengths for rule i and label j obtained from the product between the standard rule strengths and weighted rules.

5.3.5 Stage 5: Hierarchical fuzzy inference

The final stage is the development of Mamdani rule-based inference systems [215] for the different thematic subsystems. The inference systems provide assessments of driving risk using the fuzzy sets from stage 2 and the normalised weighted rules from stage 4. In the Mamdani inference systems, the output of each rule is a fuzzy set derived from the aggregation of fuzzified inputs using fuzzy operators and rule strengths obtained from the normalised weighted rules as shown by Equation 5.7. The output fuzzy sets produced by the entire set of rules are combined into a single fuzzy set using the maximum aggregation method. The resulting fuzzy set represents the likelihood of the interaction between the inputs belonging to a specific driving risk category.
For example, imagine two rules with the following modified rule strengths:

$$\begin{aligned}
&Rule1: low\ risk = 0.1, \quad moderate\ risk = 0.7, \quad high\ risk = 0.3\\
&Rule2: low\ risk = 0.2, \quad moderate\ risk = 0.2, \quad high\ risk = 0.1
\end{aligned}$$

The final decision fusion about the driving risk is:

$$\begin{aligned}
&max(0.1, 0.2) = 20\%\ likelihood\ of\ low\ risk\\
&max(0.7, 0.2) = 70\%\ likelihood\ of\ moderate\ risk\\
&max(0.3, 0.1) = 30\%\ likelihood\ of\ high\ risk
\end{aligned}$$

To compute the inference of the entire hierarchical fuzzy system, the output FSs of preceding subsystems in the hierarchy are reduced to crisp values using a defuzzification method (e.g., centroid defuzzification [348] and Mean Of Maxima (MOM) defuzzification [297]) and fed into subsequent subsystems. This is because fuzzy logic systems only accept non-fuzzy inputs, which are transformed to fuzzy sets by the fuzzification stage.

Section 5.5 presents the results of the application of SIFA in assessing HGV driving scenarios, which will elucidate the above stages.

5.4 Experimental Design

As mentioned in the motivation of this thesis (Section 1.1, page 1), HGV drivers' responses are mostly influenced by drivers' personal traits and external contextual factors. The influence of these factors are not captured in existing driver data and/or AI models; however, they can at least partially be obtained from stakeholders expertise in the HGV sector. In addition, when the influence of the factors are considered in a linear fashion as described by SIA in Section 4.5.5 (page 92), it fails to capture the realistic interactions between the factors and HGV driving risk. This section presents the experimental design of applying SIFA to capture information about the impact of the interaction between contextual factors on HGV driving risk from stakeholders in the HGV sector. The section also presents the modelling and inference techniques based on fuzzy logic adopted by SIFA for embedding and fusing contextual information in the assessment of HGV driving risk. Table 5.1 presents the number of participants recruited in different stages of the application process of SIFA. More details about the experimental design and recruitment of participants is provided below.

5.4.1 Hierarchical Structure

Since this thesis focuses on the development of an online HGV driving risk assessment framework, it only considers contextual factors that can be currently predicted using AI approaches and/or data. Using the contextual factors identified by SIA in the workshops (Section 4.5.1, page 82), the following factors are considered: driving styles, driver affective states, driver distraction, weather conditions, road types and traffic conditions.

Table 5.1: Number of participants recruited during the application process of SIFA in HGV driving risk assessment.

SIFA application process	**Number of Participants**			
	HGV drivers	**HGV Transport Managers**	**Road safety officers**	**Researchers**[b]
Questionnaire design	-	3	-	6
Questionnaire completion	28	30	29	34
Design of user stories	10	-	-	-
Evaluation of user stories	38	-	-	-

The nine stakeholders recruited by SIA to validate and update the contextual factors (i.e., a university professor in Psychology specialised in HGV driver behaviour, three HGV transport managers, and five researchers specialised in driver behaviour analysis and fuzzy logic), assisted in designing the hierarchical fuzzy system, defining the fuzzy sets and designing the questionnaire. These stakeholders grouped the factors into two themes: (1) driver-related factors, and (2) external factors, as shown in Figure 5.6. The outputs of the subsystems are combined to determine the overall driving risk. This hierarchical structure makes the system easier for humans to understand, process, evaluate and improve. Imagine a standard FLS with the six input variables and their FSs. An example rule with all six variables is “A driver is driving in an ‘undivided urban road’, driving style is ’normal’, driver is ‘distracted’ and ‘tired’, weather is ‘sunny’, and road is ‘busy’.” Reasoning and evaluating the interaction and impact of all six variable conditions on road safety would potentially be more difficult compared to splitting the interactions into smaller manageable subsystems.

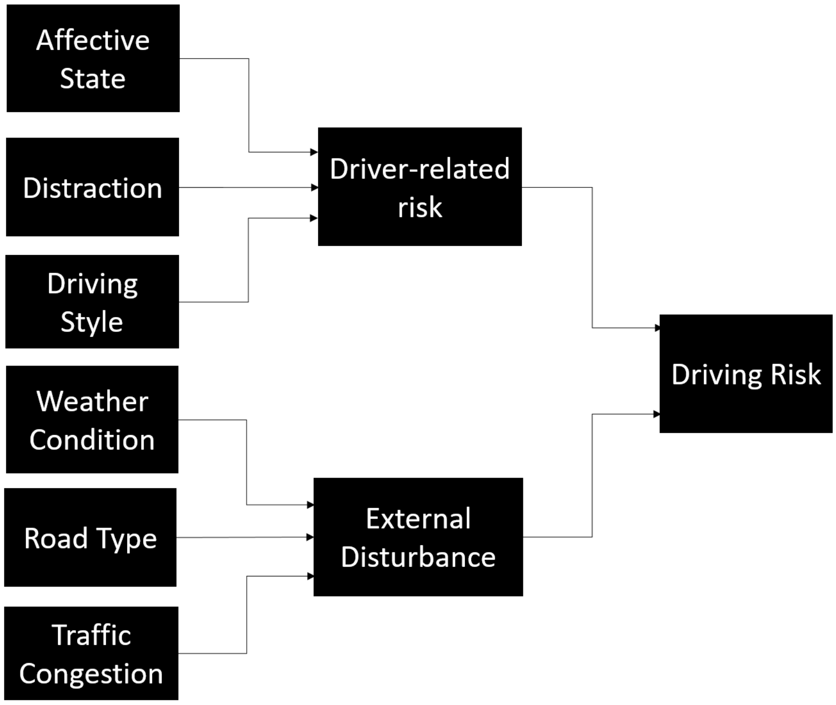


Figure 5.6: A hierarchical structure of HGV driving risk assessment developed by nine stakeholders in the HGV sector. The structure shows the grouping of HGV contextual factors into driver-related factors and external factors to enhance interpretability.

5.4.2 Fuzzy sets

The FSs defined for the different variables, intermediate features and driving risk are described below.

Subsystem 1 - Driver-related risk

The inputs of subsystem 1 are drivers' affective states, distraction and driving styles detected by intelligent driver characterisation systems. Our proposed FSs and MFs for the inputs are presented in Table 5.2. Positive affective state represents positive feelings of emotion, such as happy and energetic, while negative affective state represents negative feelings, such as angry, sad and tired. Distraction is classified into 'low', 'moderate' and 'high', where 'low' represents attentive driving and 'high' represents distracted driving. The FSs for driving styles represent the manner by which the driver operates the vehicle e.g., calm driving style means the driver is cautious and produces small number of driving incidents, while aggressive is the production of many driving incidents. Z and S-MFs are preferred for representing the extreme FSs due to their smooth transition from maximum to minimum memberships and triangular MFs are chosen for intermediate FSs to capture maximum membership at the centre. The MFs are equally distributed between 0 and 100, as shown in Figure 5.7(a).

The output of this subsystem is the driving risk associated to the co-occurrence and interaction of the driver traits. Five FSs are defined for output as shown in

Figure 5.8 i.e., 'very low', 'low', 'moderate', 'high' and 'very high'. The FSs are equally distributed between 0 and 100, where 0 is 'very low' driving risk and 100 is 'very high' driving risk.

Table 5.2: Fuzzy sets and membership functions to capture the characterisation of driver affective states, distraction and driving styles.

Driver personal traits	**Linguistic terms (fuzzy sets)**	**Membership functions**
Affective state	Negative; Neutral; Positive	Z; Triangular; S
Distraction	Low; Moderate; High	Z; Triangular; S
Driving style	Calm or Cautious; Moderate or Normal; Aggressive or Reckless	Z; Triangular; S

Subsystem 2 - External disturbance

The second subsystem takes as inputs information about environmental factors i.e., weather conditions, road types and traffic conditions. The information can also be obtained from intelligent systems or external sources, such as the Metropolitan Police UK[2]. Table 5.3 shows our proposed FSs and their respective MFs for the environmental factors. Triangular MFs are defined to model uncertainty in the prediction of weather conditions and road types, where maximum membership occurs at different class labels e.g., 1 = sunny, 2 = rainy, 3 = foggy etc. Traffic congestion is represented using Z, S and triangular MFs to capture 'low', 'moderate' and 'high' traffic characterisations. The MFs for weather conditions and road types are equally distributed between 0 and 5 for the five classes each, while the MF for traffic congestion is distributed between 0 and 100, where 0% represents 'low' traffic, 50% represents 'moderate' traffic and '100' represents 'high' traffic. Figure 5.7(b) shows the different FSs for the external factors.

The output of this system is the driving risk associated to the co-occurrence and interaction of the external factors. Similar to subsystem 1, five FSs are defined to effectively capture the uncertainty in assessing the risk, as presented in Figure 5.9, i.e., 'very low', 'low', 'moderate', 'high' and 'very high'.

Table 5.3: Fuzzy sets and membership functions to capture the characterisation of weather conditions, road types and traffic congestion.

Environmental factors	**Linguistic terms (fuzzy sets)**	**Membership functions**
Weather conditions	Sunny; Rainy; Foggy; Snowy; Icy	Triangular
Road Types	Motorway; Undivided highway Urban area; Intersection Rural area	Triangular
Traffic congestion	Low; Moderate; High	Z; Triangular; S

[2]https://www.met.police.uk/

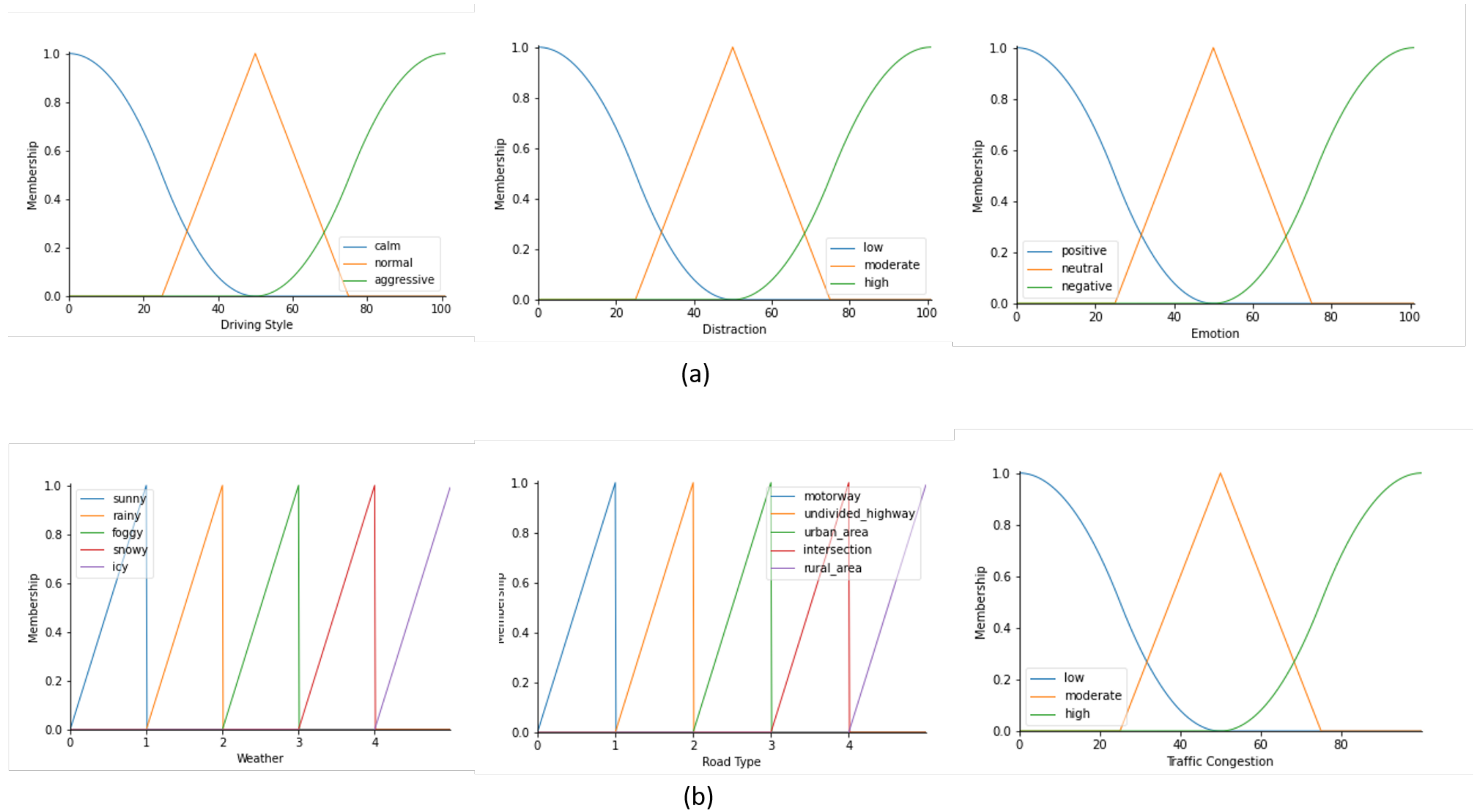


Figure 5.7: Membership functions to capture ambiguity or uncertainty in the characterisation of contextual factors. (a) Membership functions for driver traits, and (b) membership functions for external factors.

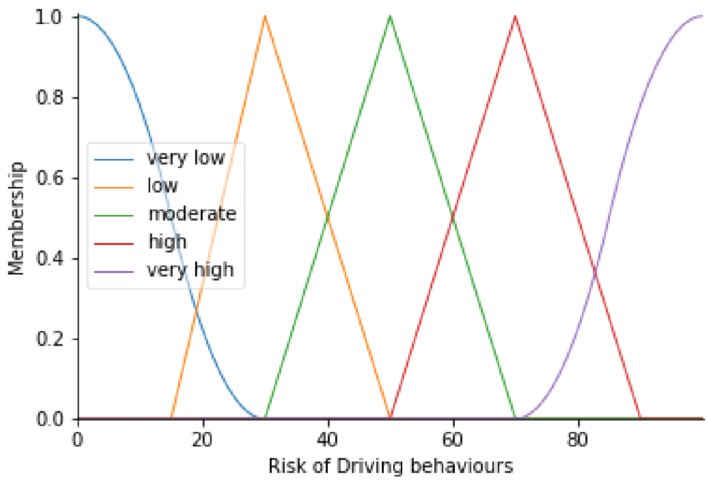


Figure 5.8: Membership function to capture ambiguity or uncertainty in the description of driving risk associated to the co-occurrence and interaction of driving behaviours.

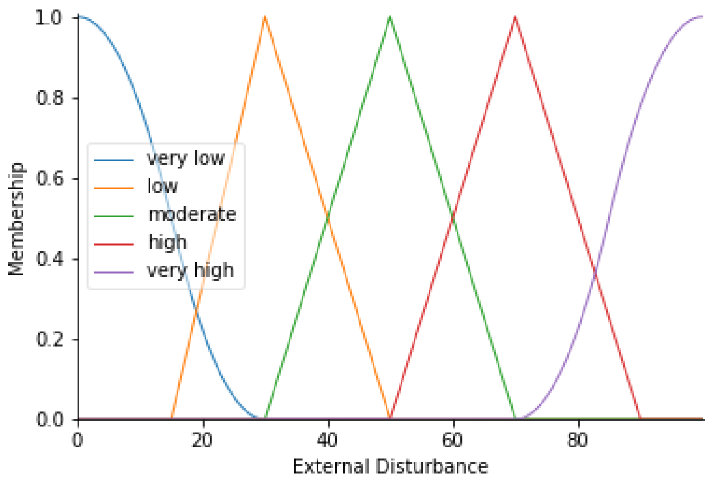


Figure 5.9: Membership functions to capture ambiguity or uncertainty in the assessment of driving risk associated to co-occurrence and interaction of external factors.

Subsystem 3 - Overall driving risk

The final subsystem combines the driving risk associated to driver traits and the driving risk associated to external factors to produce a final assessment of HGV driving risk. Three FSs are defined for the different driving risks as shown in Figure 5.10 i.e., 'low', 'moderate', and 'high'. The number of FSs were reduced from five (i.e. output MFs of subsystems 1 and 2) to three (input MFs of subsystem 3) as the stakeholders found lack of context between the linguistic terms 'very low' and 'low' or 'very high' and 'high' when considered as inputs into the system. Five FSs are defined for the final driving risk assessment as shown in Figure 5.11 i.e. 'very low', 'low', 'moderate', 'high' and 'very high'.

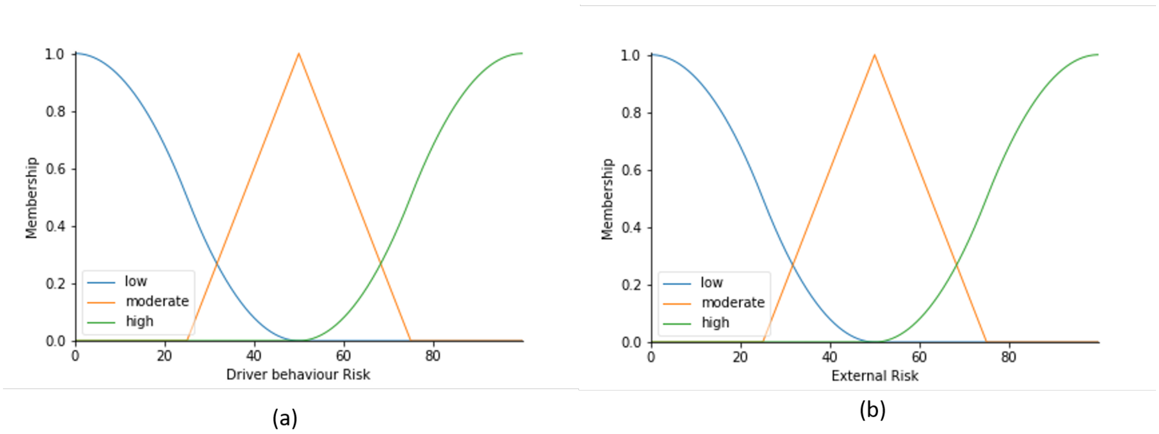


Figure 5.10: Input membership functions for determining the overall driving risk. (a) Risk associated to driver traits, and (b) risk associated to external conditions.

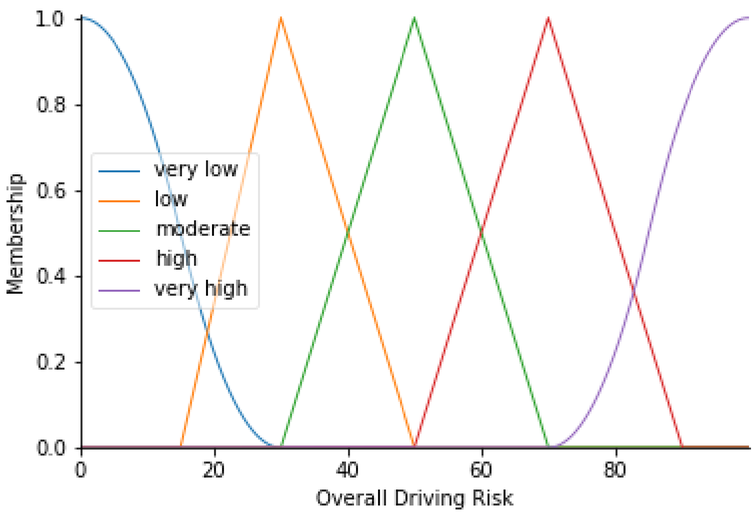


Figure 5.11: Membership function for the overall driving risk from the co-occurrence of driver traits and external factors.

The above FSs and MFs are defined by fuzzy experts based on insights from the stakeholders about the imprecision of describing and predicting risk factors. For example, the weather and road type MFs are designed to consider the imprecision of current classification machine learning models as shown in Figure 5.7. Nonetheless, the FSs and MFs are not ultimate, as other fuzzy and domain experts may have different opinions. Therefore, the FSs and MFs can be fine-tuned according to expectations, system performance and the development of new intelligent systems that characterise the risk factors.

5.4.3 Questionnaire design

To capture stakeholders' views about the combined effects and interactions of HGV contextual factors on road safety, we developed questions (found in Appendix B.1) with the interactions represented as IF-THEN statements. The questions consisted of five-point answer scales ranging from 1 meaning 'very low' risk to 5 meaning 'very high' risk, similar to the driving risk FSs described in Section 5.4.2. Figure 5.12 shows sample questions extracted from the questionnaires, asking participants to provide their opinions about the assessment of driving risk in different driving scenarios. In order to capture uncertainty in the opinions of stakeholders and ensure sufficient responses are collected, participants are allowed to select two points on the scale representing the range of certainty of their responses. The questionnaire to capture the combined impact of driver traits consisted of 27 questions representing the 27 possible combinations of the traits (subsystem 1). The questionnaire for external factors consisted of 65 questions (subsystem 2), and the questionnaire for the interaction between the risk associated to driver traits and external factors consisted of 9 questions (subsystem 3).

5.4.4 Questionnaire participant recruitment

Professionals from the four stakeholder groups were recruited to complete the questionnaires, i.e., HGV transport managers, road safety professionals, HGV drivers and researchers. These groups were established in Section 4.4.3 (page 80). Professionals were recruited by sending mass messages to individuals whose job titles and expertise matched any of the four stakeholder groups. For a wider reach, LinkedIn [328] platform, contacts in the University of Nottingham, and UK's University Transport Study Group were explored. No compensation was offered for participation. 121 participants from the UK completed the questionnaires; 28 HGV drivers, 34 researchers, 30 HGV transport managers, and 29 road safety professionals.

5.4.5 Data modelling and fuzzy system configuration

The questionnaire responses are transformed into two-element tuples consisting of two discrete points selected in the five-point rating scale. This is to ensure easy compatibility with Wagner *et al.* [19] uncertainty rule generation strategy, described in Section 5.3.4 (page 109). For certain responses (where participants select a single discrete point on the scale), the values for the elements in the tuple are same. E.g., if a participant selects only '3' on the scale, their response is represented in the tuple as (3,3).

What is the level of HGV driving risk if the following behaviours occur at the same time?

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Emotion is **positive;** Distraction is **high;** Driving style is **aggressive**	☐	☐	☐	☐	☐

(a)

What is the level of external disturbance on HGV driving if the following conditions occur at the same time?

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Weather is **foggy;** Traffic Congestion is **moderate;** Road in an **urban area**	☐	☐	☐	☐	☐

(b)

What is the overall level of driving risk considering the risk associated to driving behaviours and external disturbance?

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Driver-related risk is **high;** External disturbance is **high**	☐	☐	☐	☐	☐

(c)

Figure 5.12: Sample questions to capture the synergy of contextual factors on driving risk. (a) Driver traits; (b) external factors; and (c) interaction between the risk associated to driver traits and external factors.

For the fuzzy inference systems within the three subsystems of the hierarchy, Mamdani rule-based fuzzy inference [215] is employed as described in Section 5.3.5 (page 110). To process the outputs of subsystems 1 and 2, two popular defuzzification methods were implemented i.e., centroid defuzzification [348] and Mean Of Maxima (MOM) defuzzification [297]. Centroid defuzzification considers the centre of the area under the output fuzzy sets to produce a crisp value. Conversely, MOM computes the mean of the fuzzy set corresponding to the maximum value of the membership function.

5.4.6 Evaluation protocol

To evaluate the reliability of SIFA taking into consideration the lack of labelled HGV driving assessment data, user stories of driving scenarios were designed. The assessments of the stories computed by SIFA are compared with the assessments suggested by HGV drivers (human-in-the-loop evaluation). 10 HGV drivers were interviewed to revise and validate the user stories encompassing realistic HGV driving scenarios with different interactions between driver traits and external conditions. The user stories consist of the contextual factors established in the first stage of SIFA for online HGV driving assessment (Figure 5.6, page 113) i.e., weather, type of road, traffic, driving style, driver attention/distraction, driver affective state (e.g. emotion, fatigue). We developed the scenarios based on the following research questions: is driving risk affected by external disturbances, even if a driver's actions and behaviours are good? and to what extent do risky external conditions and driver traits affect road safety? The initial stories were developed with the help of the stakeholders recruited in stage 1 of SIA and they are found in Appendix B.2. A sample initial user story is *" A HGV driver is travelling on a sunny motorway with about 50 vehicles per lane of the road. The driver has a total of 10 seconds of driving at the maximum speed limit. The driver is attentive and well rested."*

During the interviews with the drivers, the initial user stories are presented to them and they are asked the following questions:

- On a scale from 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?
- If the driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?
- How will you improve the wordings of the user story to make it more understandable to drivers?

The drivers suggested that describing traffic congestion in terms of the pace of vehicle movement is more understandable than in terms of the number of vehicles travelling at a particular time. In addition, in their opinion it is easier to interpret driving styles using the number of driving incidents produced per mile rather than for an entire journey. Therefore, the driving scenarios were considered per mile of the road. For instance, the sample user story presented above was revised to *"A HGV driver is travelling on a sunny motorway with traffic moving at a fast pace. The driver has a total of 10 seconds of driving at the maximum speed limit within 1 mile of the road. The driver is attentive and well rested."*

The revised user stories (found in Table 5.4) were distributed to 38 HGV drivers in the UK recruited via LinkedIn. The first user story was used as a baseline with safe driving traits and favourable external conditions i.e., "A HGV driver is travelling on a sunny motorway with traffic moving at a fast pace. The driver maintains a speed below the maximum speed limit. The driver is attentive and well rested". This baseline story was presented to the drivers as an ideal driving scenario to facilitate their understanding of the other stories. For each story, the drivers were asked the following questions:

- What level of risk do you consider the above driving scenario?
- Which contextual factors in the driving scenario were influential in your assessment?

The questionnaire presented to drivers can be found in Appendix B.3. The first question had responses from 'very low' to 'very high', similar to the output of our system. The responses for the second question were the six inputs to the system (driving styles, driver affective state, driver distraction, weather conditions, road types and traffic conditions) and the different groups of inputs (driver-related factors and external factors). The purpose of the first question was to compare drivers' responses with the risk calculated by SIFA's hierarchical fuzzy system. The purpose of the second question was to compare the drivers' reasoning about risky factors for each scenario with the risks calculated by the different subsystems.

Table 5.4: User stories developed by HGV drivers to represent realistic HGV driving scenarios

User story No.	User story description
1	A HGV driver is travelling on a sunny motorway with traffic moving at a fast pace. The driver maintains a speed below the maximum speed limit. The driver is attentive and well rested.
2	A HGV driver is approaching a busy roundabout in a foggy weather. The driver reduces their speed well below the maximum speed limit of the road. The driver is attentive and well rested.
3	A HGV driver is travelling in a rainy undivided highway with traffic moving at a fast pace. The driver has a total of 20 seconds of driving at the maximum speed limit within 1 mile of the road. The driver is well rested but distracted.
4	A HGV driver is approaching a clear roundabout in a sunny weather. The driver maintains their speed. The driver is attentive but has a negative mood.
5	A HGV driver is travelling on a sunny undivided highway with traffic moving at a moderate pace. The driver has a total of 6 harsh braking incidents within 1 mile of the road. The driver is distracted and frustrated by their in-vehicle delivery management system.
6	A HGV driver is travelling in a foggy motorway with traffic moving at a slow pace. The driver produces a total of 1 harsh braking incidents within 1 mile of the road. The driver is distracted and in a negative mood.
7	A HGV driver is travelling on a icy rural road with traffic moving at a moderate pace. The driver produces a total of 2 rash overtaking incidents within 1 mile of the road. The driver is attentive but angry due to pressure from their manager to complete a job.
8	A HGV driver is travelling in a snow urban road with traffic moving at a moderate pace. The driver reduces their speed well below the maximum speed limit and maintains an appropriate distance from preceding vehicles. The driver is attentive and well rested.
9	A HGV driver is travelling on a sunny rural road with traffic moving at a fast pace. The driver produces a total of 1 harsh acceleration incidents within 1 mile of the road. The driver is in a positive mood but distracted.
10	A HGV driver is travelling on a snowy urban road with traffic moving at a fast pace. The driver produces no incidents within 1 mile of the road. The driver is attentive but frustrated by the amount of jobs assigned to them.

In order to capture uncertainties in drivers' perceptions, a multi-select response format questionnaire was implemented i.e., they could select more than one answer for the same question. It is important to note that the performance of our framework is not presented to the drivers to avoid any bias. In addition, the drivers recruited for evaluation are different from those recruited to help co-design the system.

To ensure that the fuzzy inference system developed at the last stage of SIFA can effectively fuse information about the driving scenarios, the contextual information (i.e., weather, road type, traffic, driving style, affective state and distraction) are transformed into their respective FSs with a membership degree of 1. Table 5.5 shows the input FSs for the different user stories.

Table 5.5: Resulting linguistic terms describing the user stories in Table 5.4.

User story number	**Inputs**					
	Driving behaviours			**External disturbance**		
	Driving style	**Affective state**	**Distraction**	**Weather**	**Road Type**	**Traffic Congestion**
1	Cautious	Positive	Low	Sunny	Motorway	Low
2	Cautious	Positive	Low	Foggy	Intersection	High
3	Reckless	Positive	High	Rainy	Undivided highway	Low
4	Reckless	Negative	Low	Sunny	Intersection	Low
5	Aggressive	Negative	High	Sunny	Undivided highway	Moderate
6	Normal	Negative	High	Foggy	Motorway	High
7	Aggressive	Negative	Low	Icy	Rural road	Moderate
8	Cautious	Positive	Low	Snowy	Urban road	Moderate
9	Normal	Positive	High	Sunny	Rural road	Low
10	Cautious	Negative	Low	Snowy	Urban road	Low

The assessments of the driving scenarios provided by the 38 HGV drivers are compared with the assessments produced by SIFA using the following proposed evaluation metrics:

$$Reliability_{level-1} = \frac{Total(max_{output\,fuzzy\,set}{=}max_{prob\,category})}{Total(user\ stories)} \tag{5.8}$$

$$Reliability_{level-2} = \frac{Total(max_{prob\,category}\ \mathbf{IN}\ Top2_{output\,fuzzy\,set})}{Total(user\ stories)} \tag{5.9}$$

Equation 5.8 (known as 'level-1' reliability) measures the agreement between SIFA's output fuzzy set with maximum membership and the driving risk category with highest probability obtained from the assessments provided by the 38 HGV drivers. Equation 5.9 (known as 'level-2' reliability) captures uncertainty in the assessment of driving risk by taking into consideration the uncertainty in output fuzzy sets produced by SIFA. It calculates the agreement between the top 2 output fuzzy sets produced by SIFA and the driving risk categories with highest probability in the probability distribution plot of the assessments provided by the drivers.

Lastly, the driving risk assessments produced by SIFA are compared to the assessments produced by SIA and current intelligent HGV driving assessment (CIDA) approaches. The application of SIA and CIDA in assessing driving risk is described in Section 4.5.5 (page 92).

5.5 Results and Discussion

This section presents the analysis of the responses provided by the stakeholders about the combined effects of contextual factors on HGV driving risk. Subsequently, it discusses the evaluation of the driving risk assessments produced by SIFA for the different driving scenarios presented in Table 5.4, taking into consideration the lack of labelled HGV driving risk assessment data. Lastly, it concludes with the limitations of SIFA.

5.5.1 Analysis of stakeholder responses

After collecting the responses of 121 stakeholders about the impact of the combined effects and interaction of contextual factors on HGV driving risk, the rule weight strategy proposed by Wagner *et al.* [19] is utilised to transform the responses into weighted fuzzy rules, as discussed in Section 5.3.4. Figures 5.13, 5.14, 5.15 and 5.16 show the resulting weighted rules obtained from the stakeholders' responses. For each rule, weights are assigned to the driving risk labels: 'very low', 'low', 'moderate', 'high' and 'very high'. These weights capture and represent the variance in stakeholders' views about the interaction of factors. For example, if we consider Rule 1 in Figure 5.13 i.e., *"IF Emotion is negative And Distraction is high And Driving Style is aggressive"*, it is observed that the following weights have been assigned to the risk levels: 'very low' equal 0.00, 'low' equals 0.00, 'moderate' equals 0.05, 'high' equals 0.44 and 'very high' equals 1.00. This implies that within the uncertainty interval defined by each participant for Rule 1, the rule was considered to be between 'moderate' to 'very high' risk with the highest likelihood of 'very high' risk. Shading has been applied in the figures to highlight driving risk labels with higher weights for each rule.

Figure 5.13 represents the resulting rule weights for the interaction between driver traits. It is observed that all the rules with distraction equals 'high' or driving style equals 'aggressive' irrespective of the driver's affective state have their resulting driving risk being 'high' or 'very high' i.e., the driving risk labels with the highest weights in the rules. Whereas, some rules with driver affect equals 'negative' (i.e. rules 6, 11, 17 and 21) have their resulting driving risk as 'moderate'. This implies that high distraction and aggressive driving styles have a higher impact on driving risk, when interacting with other driver traits compared to drivers' negative affective states. When we only consider their independent influence on driving risk as analysed in Chapter 4 (Figure 4.16,page 96), it can be observed that negative affective states has a higher impact on drivers' responses, which may be highly correlated to drivers' perceived distraction state and aggressive driving styles [6]. Therefore, by considering the interactions between traits, such correlations are eliminated to produce true causal relationships and more realistic effects of driver traits. Furthermore, it is observed that the assessment of driving behaviours is definitely uncertain by nature, indicated by the different levels of uncertainty in the rules; even for rules with all driver traits being good (Rule 2) or bad (Rule 3).

Examining the weighted rules for the interaction between external factors shown in Figures 5.14 and 5.15, it is observed from participants' responses that harsh

		THEN Driving Risk				
Number	**Rules for driver behaviour risk**	very low	low	moderate	high	very high
1	IF Emotion is negative And Distraction is high And Driving Style is aggressive	0.00	0.00	0.05	0.44	1.00
2	IF Emotion is positive And Distraction is low And Driving Style is calm	1.00	0.54	0.10	0.00	0.00
3	IF Emotion is positive And Distraction is high And Driving Style is aggressive	0.00	0.06	0.28	1.00	0.87
4	IF Emotion is negative And Distraction is low And Driving Style is aggressive	0.00	0.12	0.44	1.00	0.60
5	IF Emotion is positive And Distraction is low And Driving Style is aggressive	0.00	0.15	0.74	1.00	0.33
6	IF Emotion is negative And Distraction is low And Driving Style is normal	0.12	0.48	1.00	0.17	0.00
7	IF Emotion is negative And Distraction is high And Driving Style is calm	0.00	0.30	0.62	1.00	0.28
8	IF Emotion is positive And Distraction is high And Driving Style is normal	0.00	0.19	0.97	1.00	0.30
9	IF Emotion is positive And Distraction is low And Driving Style is normal	1.00	0.55	0.35	0.02	0.00
10	IF Emotion is negative And Distraction is high And Driving Style is normal	0.00	0.13	0.57	1.00	0.24
11	IF Emotion is negative And Distraction is low And Driving Style is calm	0.15	0.62	1.00	0.02	0.00
12	IF Emotion is positive And Distraction is high And Driving Style is calm	0.00	0.33	0.85	1.00	0.08
13	IF Emotion is negative And Distraction is moderate And Driving Style is aggressive	0.00	0.02	0.33	1.00	0.65
14	IF Emotion is positive And Distraction is moderate And Driving Style is calm	0.35	0.93	1.00	0.28	0.00
15	IF Emotion is neutral And Distraction is high And Driving Style is aggressive	0.00	0.02	0.27	1.00	0.76
16	IF Emotion is neutral And Distraction is low And Driving Style is calm	1.00	0.96	0.40	0.11	0.00
17	IF Emotion is negative And Distraction is moderate And Driving Style is calm	0.14	0.84	1.00	0.30	0.00
18	IF Emotion is neutral And Distraction is moderate And Driving Style is normal	0.34	1.00	0.90	0.22	0.00
19	IF Emotion is neutral And Distraction is low And Driving Style is normal	0.94	1.00	0.36	0.00	0.04
20	IF Emotion is positive And Distraction is moderate And Driving Style is normal	0.64	1.00	0.73	0.09	0.00
21	IF Emotion is negative And Distraction is moderate And Driving Style is normal	0.02	0.62	1.00	0.29	0.00
22	IF Emotion is neutral And Distraction is high And Driving Style is normal	0.00	0.36	0.68	1.00	0.28
23	IF Emotion is neutral And Distraction is low And Driving Style is aggressive	0.00	0.09	0.66	1.00	0.39
24	IF Emotion is neutral And Distraction is high And Driving Style is calm	0.00	0.22	0.74	1.00	0.17
25	IF Emotion is neutral And Distraction is moderate And Driving Style is calm	0.36	0.82	1.00	0.31	0.00
26	IF Emotion is neutral And Distraction is moderate And Driving Style is aggressive	0.00	0.10	0.54	1.00	0.44
27	IF Emotion is positive And Distraction is moderate And Driving Style is aggressive	0.00	0.15	0.90	1.00	0.55

Figure 5.13: Rules obtained from 121 HGV professionals regarding the impact of the interaction between driving behaviours on HGV driving risk.

		THEN Driving Risk				
Number	Rules for External Disturbance	very low	low	moderate	high	very high
1	IF Weather is sunny AND Traffic is moderate AND Road is motorway	0.28	1.00	0.28	0.06	0.00
2	IF Road is icy AND Traffic is moderate AND Road is rural_area	0.00	0.00	0.90	1.00	0.90
3	IF Weather is snowy AND Traffic is moderate AND Road is urban_area	0.00	0.17	0.42	1.00	0.83
4	IF Weather is snowy AND Traffic is low AND Road is intersection	0.00	0.14	0.50	1.00	0.43
5	IF Weather is sunny AND Traffic is high AND Road is motorway	0.29	1.00	1.00	0.71	0.00
6	IF Weather is foggy AND Traffic is low AND Road is urban_area	0.00	0.13	0.67	1.00	0.13
7	IF Weather is sunny AND Traffic is moderate AND Road is undivided_highway	0.14	0.57	1.00	0.29	0.00
8	IF Road is icy AND Traffic is low AND Road is rural_area	0.00	0.00	1.00	0.56	0.19
9	IF Weather is rainy AND Traffic is low AND Road is intersection	0.00	0.40	1.00	0.47	0.00
10	IF Road is icy AND Traffic is low AND Road is motorway	0.00	0.22	1.00	0.78	0.00
11	IF Road is icy AND Traffic is high AND Road is intersection	0.00	0.13	0.13	0.60	1.00
12	IF Weather is rainy AND Traffic is high AND Road is undivided_highway	0.00	0.17	0.50	1.00	0.67
13	IF Weather is sunny AND Traffic is low AND Road is rural_area	0.29	1.00	0.35	0.00	0.00
14	IF Weather is snowy AND Traffic is low AND Road is rural_area	0.00	0.00	0.75	1.00	0.00
15	IF Weather is snowy AND Traffic is high AND Road is intersection	0.00	0.11	0.06	0.44	1.00
16	IF Weather is foggy AND Traffic is high AND Road is intersection	0.00	0.11	0.05	0.32	1.00
17	IF Weather is snowy AND Traffic is low AND Road is urban_area	0.00	0.00	0.79	1.00	0.21
18	IF Weather is rainy AND Traffic is low AND Road is undivided_highway	0.13	0.13	1.00	0.25	0.00
19	IF Road is icy AND Traffic is moderate AND Road is undivided_highway	0.00	0.00	0.13	1.00	0.73
20	IF Weather is rainy AND Traffic is high AND Road is intersection	0.00	0.25	0.75	1.00	0.33
21	IF Weather is foggy AND Traffic is moderate AND Road is urban_area	0.00	0.08	1.00	0.92	0.23
22	IF Weather is snowy AND Traffic is high AND Road is urban_area	0.00	0.17	0.42	0.75	1.00
23	IF Weather is sunny AND Traffic is low AND Road is undivided_highway	0.00	1.00	0.45	0.18	0.00
24	IF Weather is snowy AND Traffic is low AND Road is motorway	0.00	0.23	0.62	1.00	0.38
25	IF Weather is foggy AND Traffic is high AND Road is motorway	0.00	0.15	0.46	1.00	0.54
26	IF Weather is foggy AND Traffic is high AND Road is rural_area	0.00	0.20	0.90	1.00	0.70
27	IF Weather is sunny AND Traffic is moderate AND Road is urban_area	0.45	1.00	0.82	0.36	0.00
28	IF Road is icy AND Traffic is low AND Road is urban_area	0.00	0.17	1.00	0.83	0.33
29	IF Weather is rainy AND Traffic is moderate AND Road is rural_area	0.15	0.23	1.00	0.38	0.00
30	IF Weather is rainy AND Traffic is high AND Road is rural_area	0.00	0.58	0.75	1.00	0.17
31	IF Weather is sunny AND Traffic is low AND Road is intersection	0.29	1.00	0.43	0.36	0.00
32	IF Weather is sunny AND Traffic is high AND Road is undivided_highway	0.00	0.50	1.00	0.25	0.00

Figure 5.14: Rules obtained from 121 HGV professionals regarding the impact of the interaction between external factors on HGV driving risk.

33	IF Weather is foggy AND Traffic is low AND Road is undivided_highway	0.00	0.00	0.71	1.00	0.29
34	IF Road is icy AND Traffic is high AND Road is undivided_highway	0.00	0.20	0.90	0.70	1.00
35	IF Weather is foggy AND Traffic is moderate AND Road is motorway	0.00	0.00	0.60	1.00	0.27
36	IF Weather is snowy AND Traffic is moderate AND Road is motorway	0.00	0.20	0.60	1.00	1.00
37	IF Road is icy AND Traffic is moderate AND Road is intersection	0.00	0.00	0.33	1.00	0.60
38	IF Road is icy AND Traffic is high AND Road is rural_area	0.00	0.20	0.80	1.00	0.90
39	IF Weather is rainy AND Traffic is low AND Road is rural_area	0.00	0.33	1.00	0.42	0.17
40	IF Weather is rainy AND Traffic is moderate AND Road is undivided_highway	0.00	0.10	1.00	0.90	0.30
41	IF Weather is foggy AND Traffic is low AND Road is rural_area	0.00	0.00	1.00	0.71	0.29
42	IF Weather is sunny AND Traffic is moderate AND Road is intersection	0.10	1.00	0.70	0.30	0.00
43	IF Weather is snowy AND Traffic is high AND Road is motorway	0.00	0.14	0.29	1.00	0.57
44	IF Weather is snowy AND Traffic is moderate AND Road is rural_area	0.00	0.11	0.33	1.00	0.11
45	IF Weather is sunny AND Traffic is high AND Road is rural_area	0.00	1.00	0.88	0.25	0.25
46	IF Weather is foggy AND Traffic is moderate AND Road is rural_area	0.00	0.31	0.44	1.00	0.06
47	IF Weather is foggy AND Traffic is moderate AND Road is intersection	0.00	0.00	0.58	1.00	0.75
48	IF Weather is sunny AND Traffic is low AND Road is motorway	0.92	1.00	0.23	0.00	0.00
49	IF Weather is snowy AND Traffic is low AND Road is undivided_highway	0.00	0.09	1.00	1.00	0.45
50	IF Weather is snowy AND Traffic is moderate AND Road is intersection	0.00	0.00	0.24	1.00	0.41
51	IF Weather is sunny AND Traffic is moderate AND Road is rural_area	0.67	1.00	1.00	0.44	0.00
52	IF Weather is rainy AND Traffic is low AND Road is urban_area	0.15	0.77	1.00	0.23	0.00
53	IF Road is icy AND Traffic is moderate AND Road is motorway	0.00	0.00	0.35	1.00	0.29
54	IF Road is icy AND Traffic is moderate AND Road is urban_area	0.00	0.27	0.55	1.00	0.73
55	IF Weather is foggy AND Traffic is high AND Road is urban_area	0.06	0.06		0.53	1.00
56	IF Weather is snowy AND Traffic is high AND Road is rural_area	0.00	0.15	0.15	0.10	1.00
57	IF Weather is rainy AND Traffic is high AND Road is urban_area	0.00	0.10	1.00	1.00	0.70
58	IF Weather is foggy AND Traffic is moderate AND Road is undivided_highway	0.00	0.00	0.39	1.00	0.22
59	IF Weather is snowy AND Traffic is moderate AND Road is undivided_highway	0.00	0.00	0.25	1.00	0.50
60	IF Weather is snowy AND Traffic is high AND Road is undivided_highway	0.00	0.00	0.04	0.17	1.00
61	IF Weather is sunny AND Traffic is high AND Road is intersection	0.00	0.42	1.00	0.83	0.08
62	IF Weather is sunny AND Traffic is low AND Road is urban_area	0.58	0.75	1.00	0.00	0.00
63	IF Road is icy AND Traffic is high AND Road is urban_area	0.00	0.05	0.15	0.20	1.00
64	IF Weather is rainy AND Traffic is moderate AND Road is urban_area	0.00	0.19	1.00	0.56	0.00
65	IF Weather is foggy AND Traffic is high AND Road is undivided_highway	0.00	0.06	0.12	0.59	1.00

Figure 5.15: Rules obtained from 121 HGV professionals regarding the impact of the interaction between external factors on HGV driving risk (continuation of Figure 5.14).

weather and road conditions (i.e., snowy, icy and foggy weather) are the most risky external factors with 'high' and 'very high' driving risk irrespective of other external conditions. This is in conformity with results obtained from the literature of post-hoc driving risk analysis [235, 349, 350]. Furthermore, we observe 'low' to 'moderate' driving risk during sunny weather conditions due to better visibility compared to other weather conditions. For road types, it is noticed that driving risk is higher at intersections compared to other road types, irrespective of other external factors.

		THEN Driving Risk is				
	IF	very low	low	moderate	high	very high
1	Driving risk is moderate **And** External risk is moderate	0.00	0.08	1.00	0.08	0.00
2	Driving risk is high **And** External risk is moderate	0.00	0.00	0.27	1.00	0.00
3	Driving risk is low **And** External risk is low	0.85	1.00	0.31	0.00	0.00
4	Driving risk is high **And** External risk is high	0.00	0.00	0.00	0.40	1.00
5	Driving risk is low **And** External risk is high	0.00	0.15	0.85	1.00	0.15
6	Driving risk is moderate **And** External risk is high	0.00	0.00	0.41	1.00	0.24
7	Driving risk is low **And** External risk is moderate	0.00	0.73	1.00	0.13	0.00
8	Driving risk is moderate **And** External risk is low	0.00	0.37	1.00	0.11	0.00
9	Driving risk is high **And** External risk is low	0.00	0.00	0.44	1.00	0.11

Figure 5.16: Rules obtained from 121 HGV professionals regarding the impact of the interaction between the risk associated to driving behaviours and external factors.

Figure 5.16 shows the weights for the rules that capture the interaction between the risk associated to driver traits and the risk associated to environmental conditions. Risky environmental conditions have a high impact on road safety even when the risk associated to driving behaviours is 'low' i.e., rule 5. In addition, the responses revealed that the risk associated to environmental conditions has a greater impact on the overall driving risk compared to the risk associated to driving behaviours. This is indicated by higher weights in 'very high' driving risk for 'high' risk associated to environmental conditions (rules 5 and 6) compared to 'high' risk associated to driver traits (rules 2 and 9). These findings illustrate the importance of incorporating the impact of external conditions into the assessment of driving risk and answers the research questions:is driving risk affected by external disturbances, even if a driver's actions and behaviours are good? and to what extent do risky external conditions and driver traits affect road safety? It is important to note that these findings are uncovered as a result of the hierarchical classification of the driving environment into driver-related and environmental risks.

Lastly, it is observed that the driving risk associated to the safest driving rule (rule 3) with 'low' driver-related and external risks is considered more of 'low' than 'very low'. Whereas the driving risk associated to the most dangerous driving rule (rule 4) is considered by majority of the stakeholders to be 'very high'. This shows that stakeholders support the risk homeostasis theory of driver behaviour presented in Section 2.2.1 (page 8), which states that there is always some level of driving risk.

5.5.2 Performance evaluation of defuzzification methods

The last stage of SIFA is the development of a hierarchical Mamdani inference system using the weighted rules discussed in the previous section and the FSs developed in

Table 5.6: Defuzzified values of the assessments produced by the driver-related risk and external disturbance subsystems for the user stories in Table 5.4 using centroid and mean of maxima defuzzification methods.

	Centroid defuzzification		**Mean of Maxima defuzzification**	
User story	**Driver-related risk**	**External disturbance**	**Driver-related risk**	**External disturbance**
1	**22.24 (low)**	26.36 (low)	**6.00 (very low)**	30.80 (low)
2	**22.24 (low)**	**76.59 (high)**	**6.00 (very low)**	**93.99 (very high)**
3	71.91 (high)	50.23 (moderate)	69.99 (high)	50.00 (moderate)
4	67.11 (high)	**41.74 (moderate)**	69.99 (high)	**30.80 (low)**
5	**79.67 (high)**	45.97 (moderate)	**93.99 (very high)**	50.00 (moderate)
6	62.39 (high)	65.84 (high)	69.99 (high)	70.0 (high)
7	67.11 (high)	68.37 (high)	69.99 (high)	70.0 (high)
8	**22.24 (low)**	68.20 (high)	**6.00 (very low)**	70.0 (high)
9	**60.37 (moderate)**	33.73 (low)	**69.99 (high)**	30.80 (low)
10	41.77 (moderate)	62.64 (high)	50.00 (moderate)	70.0 (high)

Section 5.4.2. Before discussing the evaluation of the overall fuzzy inference system is performed, the performance between the different defuzzification methods are shown i.e., centroid and MOM defuzzification methods.

Table 5.6 represents the risks produced by driver-related factors and external factors of SIFA's hierarchical fuzzy Mamdani inference system for the user stories. The results show defuzzified values computed using centroid and MOM defuzzification methods. The results that differ in their driving risk category (i.e., the output FSs with maximum membership) between the methods for the different user stories are indicated in bold. For example, the driver-related risk produced by centroid defuzzification is 'low' for user story 1, while that produced by MOM is 'very low'; therefore, indicated in bold. While the driver-related risk produced by centroid defuzzification is 'high' for user story 3 and similar to the driver-related risk produced by MOM, which is also 'high'.

It can be observed that the methods produce different risk categories for driver-related risk in five user stories (i.e., user stories 1, 2, 5, 8 and 9), and two user stories for external disturbance (i.e., user stories 2 and 9). It can also be observed that MOM defuzzification produces the same defuzzified values for the same categories as it calculates the mean of the fuzzy set with maximum membership. This implies more contextual information is lost with MOM, as it ignores the output uncertainties. For example, MOM produces the same defuzzified driver-related risks for user stories 3 and 9, as shown in Figure 5.17(a). However, it is observed that user story 3 has a higher likelihood (0.6 membership degree) of being classified as 'very high' risk compared to user story 9 (0.2 membership degree). This information is totally lost with MOM compared to centroid defuzzification, as shown by the defuzzified driver-related risks produced by centroid defuzzification in Figure 5.17. Centroid defuzzification loses less information compared to MOM, and therefore seems to be a more sensible approach for transforming information from earlier subsystems in the hierarchy.

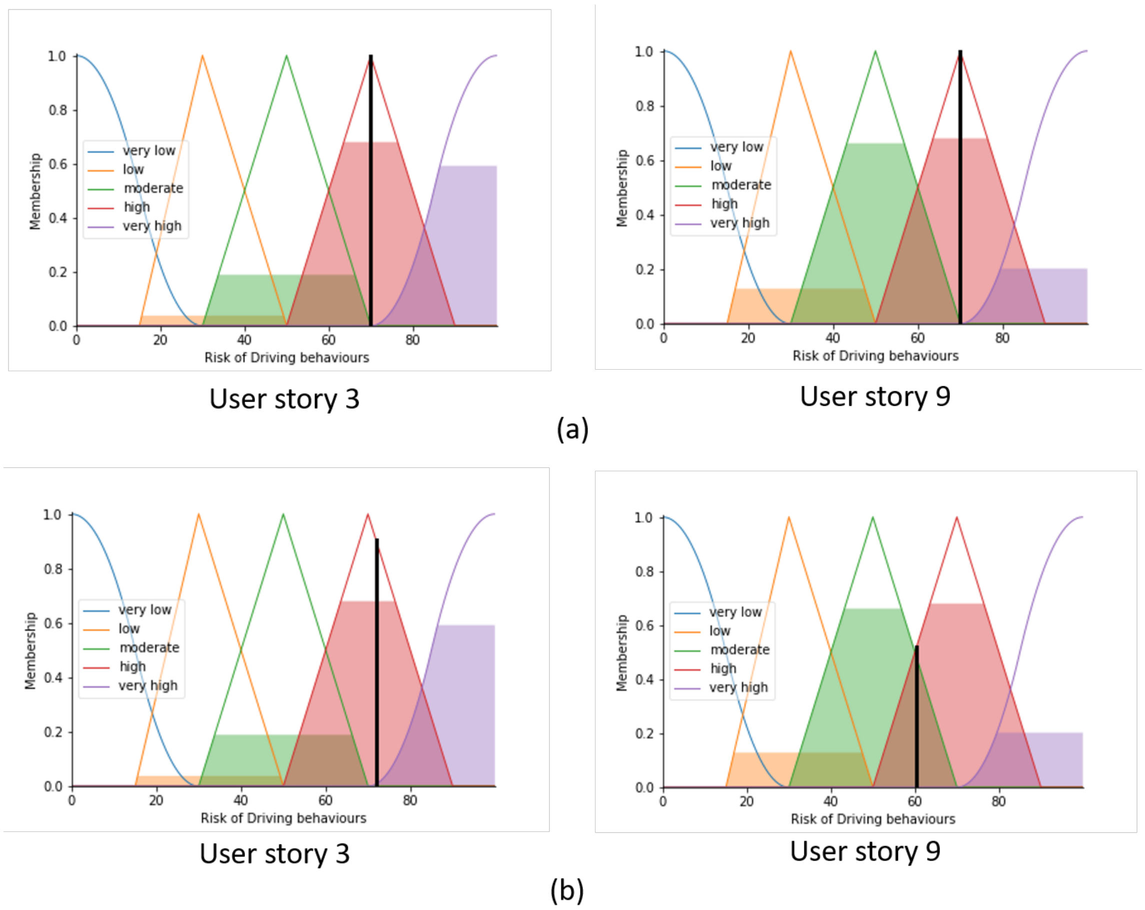


Figure 5.17: Output fuzzy sets and defuzzified values (black vertical lines) for: (a) Mean of maxima defuzzification, and (b) centroid defuzzification.

5.5.3 Evaluation of SIFA

(a) Comparison with assessments from HGV drivers

The results of SIFA and the responses of the 38 HGV drivers are presented in Figures 5.18 and 5.19. The first column represents user story numbers with their descriptions in Table 5.4 (page 5.4). The second and third columns represent the final output fuzzy sets from SIFA with centroid and MOM defuzzification methods, respectively. The shaded areas in the MFs represent the likelihood of a user story being classified as a particular driving risk category. For example, the centroid result in row 1 of Figure 5.18 means user story 2 has a higher likelihood of being classified as 'high' risk (with 0.5 membership degree), followed by 'moderate' risk (with 0.4 membership degree), a similar likelihood of being classified 'low' or 'very high' risk (with 0.1 membership degree), and a 0% likelihood of being classified as 'very low' risk. The last column represents the probability distribution of driving risk categories generated from the responses of the drivers for each user story.

To evaluate the reliability of SIFA, Equations 5.8 and 5.9 are employed as described in Section 5.4.6. Table 5.7 shows the results obtained using our proposed evaluation metrics. It can be observed that centroid SIFA has a higher level-1 re-

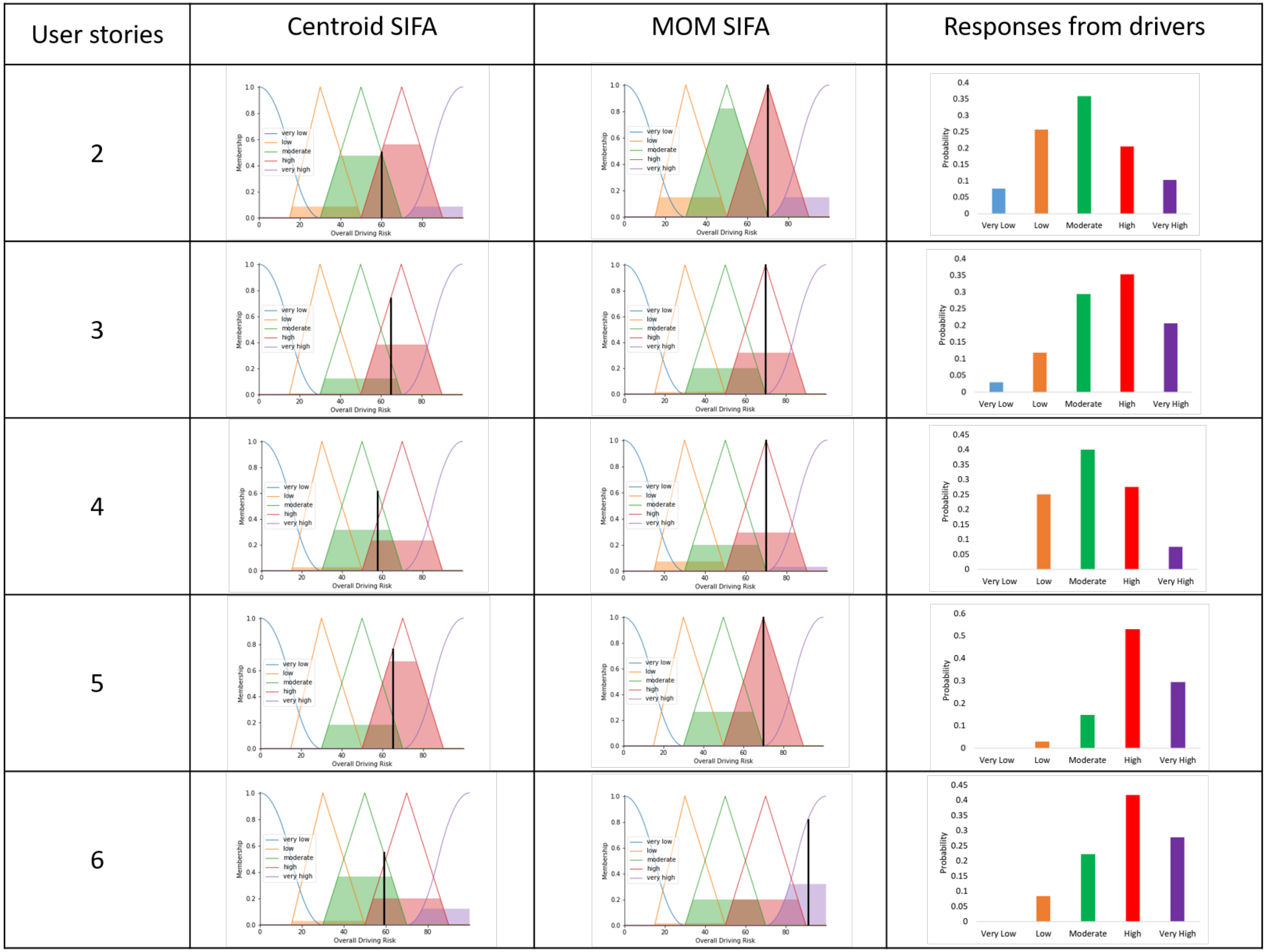


Figure 5.18: Comparison of SIFA's assessments of user stories 2 to 6 with the probability distribution generated from the responses of 38 HGV drivers.

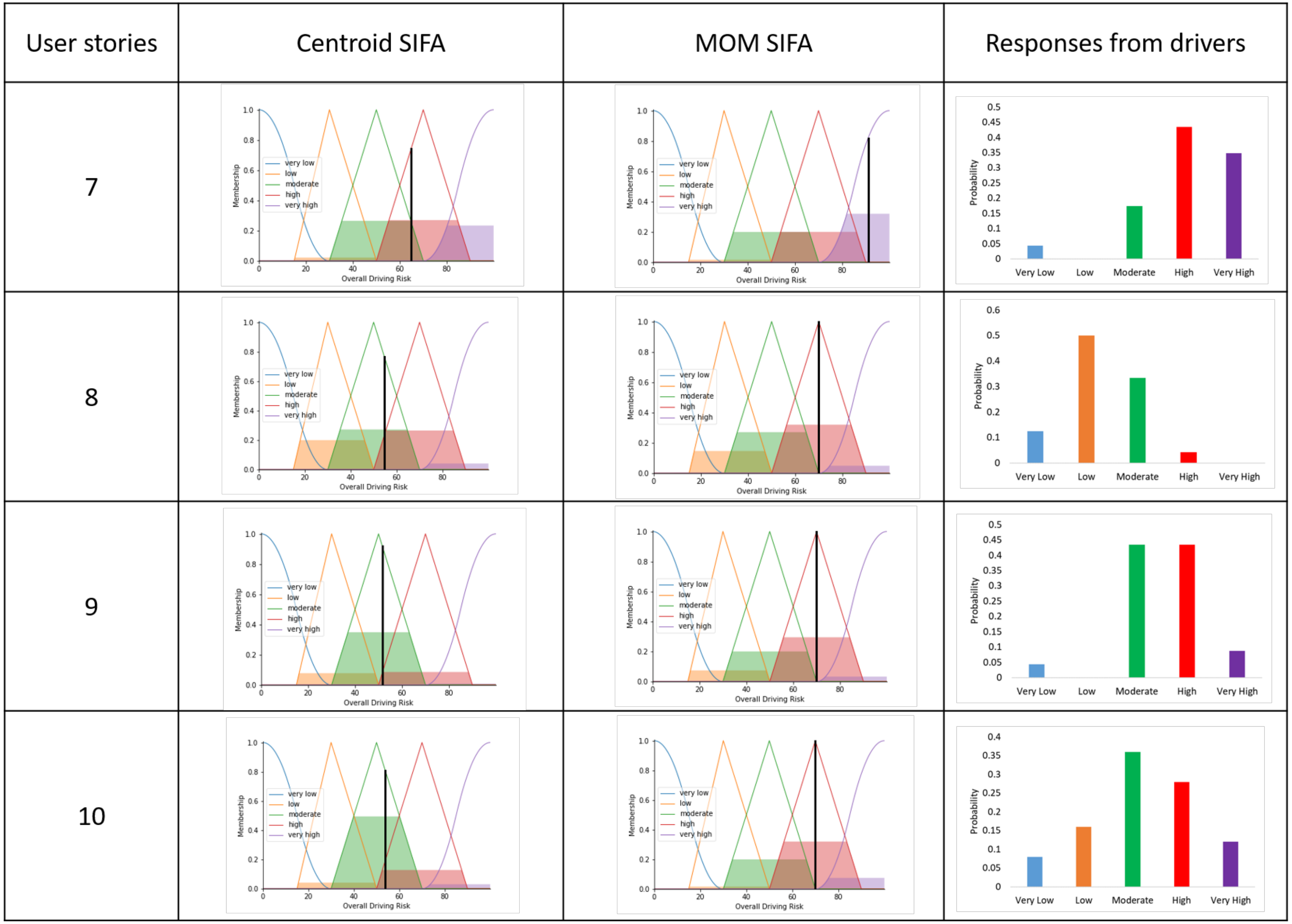


Figure 5.19: Comparison of SIFA's assessments of user stories 7 to 10 with the probability distribution generated from the responses of 38 HGV drivers.

liability (66.7%) compared to MOM SIFA (33.3%) because centroid defuzzification captures some uncertainty in intermediary subsystems, as discussed in Section 5.5.2. The 66.7% reliability of the centroid SIFA corresponds to the agreement between the maximum output FSs and the driving risk categories with highest probabilities i.e., user stories 3, 4, 5, 7, 9 and 10. While 33.3% reliability of MOM SIFA corresponds to the agreement in user stories 3, 5 and 9. When we consider the uncertainties in driving risk assessment as described in Section 5.1 (page 102), and compare the top two output fuzzy sets with highest membership degrees with the driving risk categories having the highest probabilities, centroid SIFA produces 100% level-2 reliability while MOM SIFA produces 88.9%. These results suggest that SIFA agrees totally with the assessments provided by the HGV drivers when variability in experts' subjective views and uncertainty in contextual information are considered.

Table 5.7: Agreement between the assessments produced by SIFA and the assessments provided by 38 HGV drivers regarding nine realistic HGV driving scenarios.

SIFA	Reliability (%)	
	Level-1	**Level-2**
Centroid	66.7	100
MOM	33.3	88.9

Figure 5.20 shows the responses from the drivers about the most influential factors in their assessments. We compare the frequency of factors selected as most influential with the decisions produced by subsystems 1 and 2 found in Table 5.6 i.e., risks associated to driver traits and external factors, respectively. If the drivers' responses are consistent with the assessments produced by SIFA, it implies a greater chance of trust and acceptance in SIFA. Otherwise, more investigation is required to identify the reason for the inconsistency and improve the framework or drivers' perceptions accordingly. It can be observed that majority of the drivers considered all the factors to be influential in the assessment of user story 2 and selected external factors as most influential in the overall driving risk assessment. This conforms with the decisions produced by SIFA, where the risk associated to external factors is considered 'high'. For user story 3, the drivers considered external factors to be more influential than driving behaviours, which contradicts the decisions from SIFA i.e., higher risk associated to driving behaviours than external factors. However, when the independent effects of factors are examined, the drivers selected driver distraction as the most influential factor in their assessment.

Drivers selected driver traits as the most influential in assessing user stories 4 and 5, which is in agreement with the decisions produced by SIFA i.e., higher risk associated to driving behaviours than external factors. Subsequently, the drivers consider both driving behaviours and external factors to be influential in assessing user stories 6 and 7, which is similar to SIFA's decisions. The framework produced 'low' driver-related risk for user story 8 due to the driver's cautious driving style. This is observed by majority of drivers selecting driving style as the most influential factor in their assessment of user story 8. We also observe concordance between the SIFA's decisions and drivers' assessments for user stories 9 and 10, with driver traits being more influential in user story 9 and external factors in user story 10. Regardless of the most influential factor selected, the drivers generally considered

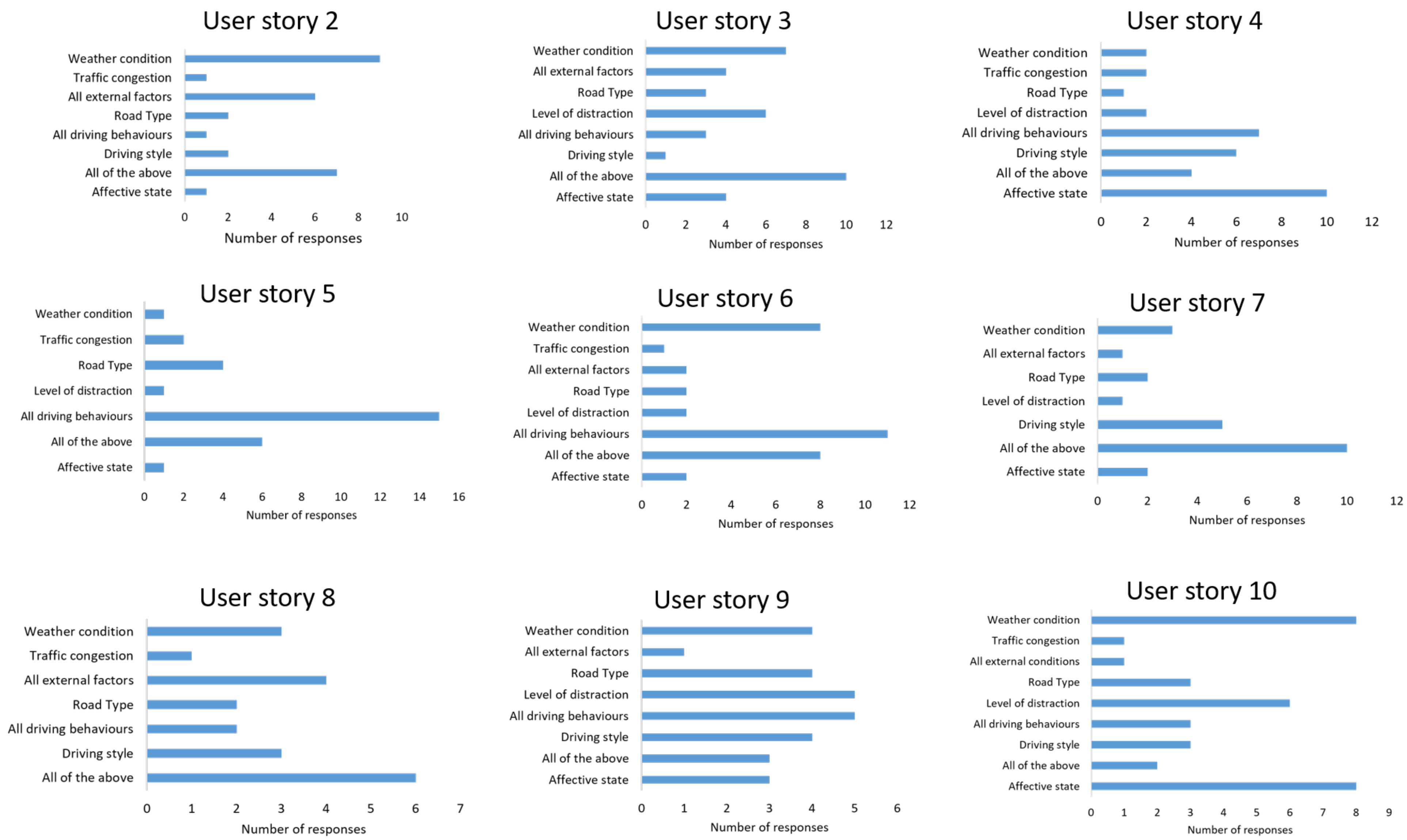


Figure 5.20: Barplots showing the influential factors suggested by 38 HGV drivers in the assessment of the HGV driving scenarios.

all contextual factors influential in their assessment of driving risk, which validates the importance of the contextual factors identified by the stakeholders for assessing driving risk.

(b) Comparison with SIA and CIDA approaches

This section compares the assessments produced by SIFA with those produced by SIA and CIDA approaches. The application of SIA and CIDA in assessing driving risk is described in Section 4.5.5 (page 92). As mentioned in Section 4.1 (page 72), CIDA approaches (e.g., [15]) do not consider the influence of contextual factors on drivers' responses and only consider the manner by which drivers operate their vehicles i.e., driving styles.

Table 5.8 shows the results of SIFA, SIA, CIDA and the assessments provided by the 38 HGV drivers (DA) for the different user stories. SIFA's results (columns 4 and 5) represent the top 2 output fuzzy sets after applying MOM and centroid defuzzification methods to transform the output fuzzy sets of subsystems 1 and 2 in the hierarchical fuzzy inference system. DA's results represent the top 2 driving risk categories with highest probabilities in the probability distribution of drivers' responses. The results that agree with the driving risk categories having highest probability in DA are indicated in bold. To obtain the results for SIA and CIDA, the normalised defuzzified contextual information obtained from stakeholders is utilised as described in Section 4.5.5 (page 4.5.5). For example, if we consider user story 1, *"A HGV driver is travelling on a sunny motorway with traffic moving at a fast pace. The driver maintains a speed below the maximum speed limit. The driver is attentive and well rested"*, the following normalised defuzzified contextual information are used:

- norm(cautious driving=7.5)$_{[0.5,1.5]}$ = 1.315
- norm(positive affect=6.8)$_{[0.5,1.5]}$ = 1.225
- norm(attentive=7.8)$_{[0.5,1.5]}$ = 1.350
- norm(sunny=5.8)$_{[0.5,1.5]}$ = 1.095
- norm(motorway=6.2)$_{[0.5,1.5]}$ = 1.150
- norm(low traffic=6.4)$_{[0.5,1.5]}$ = 1.172

Journeys are assumed to start with a risk level of 50% (where 0% represents 'very low' risk, 50% represents 'moderate' risk, and 100% represents 'very high' risk), and use Equation 4.5 (page 4.5) to compute the level of driving risk for CIDA and SIA. For CIDA, the driving risk depends only on the driver's driving style. While SIA uses all the contextual information in a linear fashion. The resulting driving risk assessments of CIDA and SIA for user story 1 are:

- $CIDA_{userstory1} = 50\% * \frac{1}{1.315} = 38.0\%$
- $SIA_{userstory1} = 50\% * \frac{1}{1.315*1.225*1.35*1.095*1.150*1.172} = 15.6\%$

Table 5.8: Comparison amongst the results produced by SIFA, SIA and CIDA approaches for the different user stories. Results that agree with the driving risk categories having highest probability in the assessments provided by drivers (DA) are indicated in bold.

User story	**CIDA**[a]	**SIA**[a]	**SIFA MOM**[b]	**SIFA Centroid**[b]	**DA**[c]
1	38.0 (low)	15.6 (very low)	very low/0.25 - low/0.3	very low/0.39 - low/0.45	-
2	38.0 (low)	**48.5 (moderate)**	**moderate/0.80** - high/1.00	**moderate/0.50** - high/0.58	low/0.25 - **moderate/0.35**
3	81.6 (very high)	90.6 (very high)	moderate/0.20 - **high/0.38**	moderate/0.10 - **high/0.40**	moderate/0.30 - **high/0.35**
4	81.6 (very high)	**59.9 (moderate)**	**moderate/0.20** - high/0.30	**moderate/0.30** - high/0.20	**moderate/0.40** - high/0.30
5	**75.9 (high)**	100 (very high)	moderate/0.25 - **high/1.00**	moderate/0.19 - **high/0.69**	**high/0.55** - very high/0.3
6	50.6 (moderate)	100 (very high)	**high/0.20** - very high/0.37	moderate/0.38 - **high/0.20**	**high/0.40** - very high/0.25
7	**75.9 (high)**	100 (very high)	**high/0.20** - very high/0.37	moderate/**high/0.25** - very high/0.20	**high/0.45** - very high/0.35
8	**38.0 (low)**	**33.5 (low)**	moderate/0.30 - high/0.35	**low/0.2** - moderate/high/0.30	**low/0.50** - moderate/0.35
9	**50.6 (moderate)**	**56.6 (moderate)**	**moderate/0.20 - high/0.30**	**moderate/0.40 - high/0.10**	**moderate/high/0.45** - very high/0.1
10	38.0 (low)	**53.3 (moderate)**	**moderate/0.20** - high/0.35	**moderate/0.50** - high/0.10	**moderate/0.35** - high/0.25

a: very low = 0 - 20, low = 21 - 40, moderate = 41 - 60, high = 61 - 80, very high = 81 -100.

b: top 2 output fuzzy sets with highest membership degrees; risk category/membership degree.

c: top 2 driving risk categories with highest probabilities; risk category/probability.

The results of SIA and CIDA are classified into a driving risk category, as follows: 'very low' = 0 - 20, 'low' = 21 - 40, 'moderate' = 41 - 60, 'high' = 61 - 80, 'very high' = 81 -100.

It can be observed that CIDA produces the same level of driving risk for user stories with similar manner of driving (i.e., driving style), even though their external conditions are different. For instance, in user stories 6 and 9, the drivers have the same driving style (i.e., number of incidents within 1 mile); however, the foggy weather and high traffic conditions of user story 6 should make the driving scenario more risky than user story 9 with sunny weather and low traffic conditions. Similarly, in user stories 1 and 2 with similar cautious driving styles (i.e., drivers maintain speeds below the maximum speed limit) but different external conditions, CIDA approaches produce the same driving risk level even though the foggy weather and busy road conditions in user story 2 should be more risky.

When the combined effects of external factors are considered in SIA, we observe fairer and more realistic driving risk levels for those scenarios. SIA's assessments show that user story 6 is more risky (100%) than user story 9 (56.6%) due to the influence of foggy weather and high traffic conditions, and user story 2 is more risky (48.5%) than user story 1 (15.6%) as a result of the foggy weather and busy road conditions. However, SIA does not capture the interactions between factors and fuses the contextual information in a linear fashion. This approach could potentially lead to an exponential increase or decrease in the overall impact on driving risk; thereby, producing severe assessments. For example, we observe such effects in the severe assessments (above 90%) produced by SIA for user stories 3, 5, 6 and 7. SIFA resolves SIA's issues of highly correlated factors and exponential effects by capturing the interactive effects of the factors as observed in SIFA's more realistic assessments of user stories 3, 5, 6 and 7.

Furthermore, SIA and CIDA fail to capture and represent the uncertainties and imprecision in HGV driving assessment, as observed by their crisp results. HGV driving assessment is uncertain by nature due to differences in the interpretations of driver traits, imprecision in the definition of driving behaviours, uncertainties produced by sensor readings, and most importantly, differences in experts' subjective views and opinions. These uncertainties can be clearly observed in SIFA's output fuzzy sets with their degrees of membership between 0 and 1 and the driver's responses (i.e., probability distribution plots) shown in Figures 5.18 and 5.19. It can be observed that even when we consider the interval between highest probabilities of drivers' response, as shown in column 'DA' in Table 5.8, CIDA does not agree that much with the drivers' responses (about 55.5% agreement). The agreement increases with SIA (about 88.8% agreement) due to the involvement of stakeholders and contextual factors in the decision-making process. We observe highest agreement between the outputs of SIFA (fuzzy sets or intervals) and drivers' responses, even when the highest probability of drivers' responses are considered only. The agreement between SIFA and DA is due to stakeholders' support and synergy between factors. SIFA with centroid defuzzification shows greater agreement than SIFA with MOM defuzzification due to the greater loss of information with MOM as shown in Figure 5.17 (page 129).

5.5.4 Limitations of SIFA

This section provides some limitations of SIFA that require improvement and further analysis in the future.

- The hierarchical fuzzy Mamdani inference system can be further extended to consider additional sub-themes within the subsystems to improve performance and interpretability. For example, driver affective states and traffic congestion could form a separate subsystem in the hierarchy as its been suggested by some studies that driving in slow traffic elicits anger and stress [351, 352].

- SIFA assumes equal weights for all stakeholders and contextual factors. However, as the characteristic of HGV driving evolve e.g., introduction of new technologies, laws, and infrastructure, we may have to revisit the weights and prioritise factors with higher influence as well as stakeholders with greater insights about the domain.

- SIFA is a stakeholder-supported framework. Its reliability and effectiveness depends on the reliability of information captured from stakeholders. Therefore, the recruitment of stakeholders and the design of questionnaires (i.e., stages 1, 2, and 3 of SIFA) are vital to the success of SIFA. In the future, we plan to further improve these stages by developing a method of measuring stakeholder reliability.

- SIFA loses contextual information during the defuzzification process. In the future, we plan to reduce the loss of information by exploring or developing new defuzzification methods that retain information captured by the output fuzzy sets.

5.6 Summary

This chapter introduced a new framework called **S**takeholder-supported **I**ntelligent **F**uzzy driving **A**ssessment (SIFA) that extends the framework introduced in Chapter 4 (SIA) using a hierarchical fuzzy Mamdani inference system. SIFA complements SIA by capturing the simultaneous interactions and uncertainties between contextual factors and driving risk with the help of stakeholders and fuzzy experts. Information about the synergistic and interactive effects of the factors on driving risk are captured from key stakeholders in the HGV sector using questionnaires, aggregated and utilised to develop a hierarchical fuzzy Mamdani inference system. SIFA also models uncertainties in the driving system to produce more reliable, context-aware, comprehensible and decomposable assessments of HGV driving risk.

We apply SIFA to assess the road safety risk of HGV driving scenarios. We recruit 121 professionals in the HGV sector, i.e., 28 HGV drivers, 34 researchers, 30 HGV transport managers, and 29 road safety professionals, to capture the combined effects, interactions and uncertainties of HGV contextual factors (described by stages 1, 2 and 3 of SIFA). The interactions and effects are aggregated and embedded into the hierarchical fuzzy Mamdani inference system in stage 4 for fusing contextual information from intelligent systems. The output of the decision fusion process is the likelihood of the interaction between the inputs belonging to a specific driving risk category.

To evaluate the reliability of SIFA, 10 user stories representing realistic HGV driving scenarios are developed with the help of HGV drivers. The decision fusion results produced by SIFA for the different user stories are compared with the assessments produced by SIA, CIDA approaches, and 38 HGV drives. The results show that the decisions produced by SIFA agree with the decisions from drivers about the driving risk of the user stories, especially when we consider uncertainty in assessing driving risk and human perception. In comparison with SIA and CIDA approaches, SIFA produces more realistic and trustworthy driving risk assessments. In addition, the decomposable nature of SIFA makes its decisions easy to understand and makes the framework easy to extend in the future.

The next chapter concludes this thesis by outlining the key findings of the research and summarising the major contributions. In addition, it presents limitations of the research and potential future directions of work.

Chapter 6

Conclusion

The initial aim of this research was to develop and evaluate an end-to-end context-aware intelligent system that provides online assessments of heavy goods vehicle (HGV) driving risk using multi-modal data streams of driving behaviours and external factors. However, due to the absence of multi-modal data streams that simultaneously capture driving behaviours and external factors, this thesis focused on a proof-of-concept implementation of the system using available data sources and insights from domain experts. The thesis had the following aims: (1) develop and evaluate new AI models that can accurately and privately detect HGV driving behaviours, and (2) develop and evaluate an intelligent system that can automatically assess the impact of detected driving behaviours on road safety taking into consideration their synergy with external conditions. The first aim has been realised by introducing novel AI techniques in Chapter 3 to improve the accuracy of detecting driving behaviours and protecting the identities of drivers when processing facial images. The second aim has been achieved by introducing a novel stakeholder-supported intelligent fuzzy driving assessment (SIFA) framework in Chapter 5, which is an extension of the stakeholder-supported intelligent driving assessment (SIA) framework introduced in Chapter 4.

The remainder of this chapter summarises the contributions of this thesis, and links the contributions to the research questions established in the beginning of the thesis (Section 1.2, page 1.2).

6.1 Contributions

The main contributions of this thesis are summarised in Figure 6.1. The figure represents our proposed end-to-end online context-aware driving assessment framework that automatically analyses the road safety risk of a driver's actions and behaviours by taking into account their interactions with external conditions. The following paragraphs describe the different contributions.

1. A novel intelligent multifaceted driver characterisation framework (Chapter 3). The framework consists of novel artificial intelligence (AI) models that process multi-modal data to produce more accurate, reliable and privacy-preserving predictions of driving styles, distracted postures and affective states. The

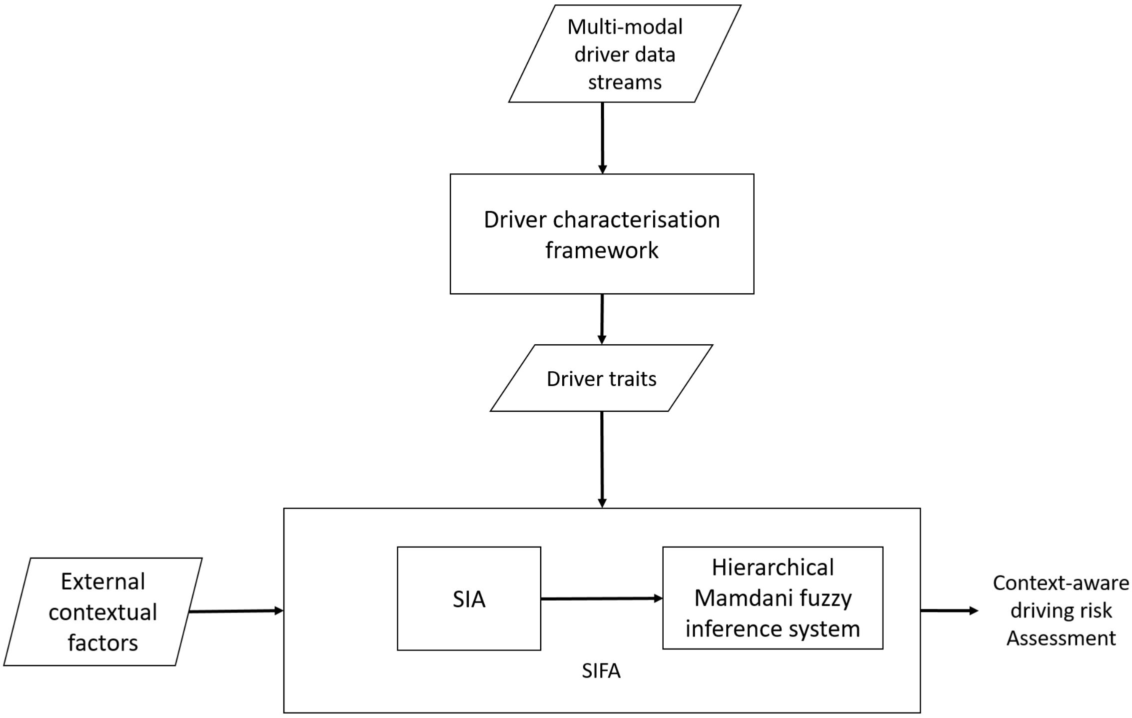


Figure 6.1: Our proposed intelligent context-aware HGV driving risk assessment framework

development of this framework answers the first and second research questions of this thesis i.e., "how can the accuracy and reliability of detecting HGV driver behaviour be improved taking into consideration its multifaceted nature?" and "how can the privacy of drivers be protected when processing data that can easily expose their identities, such as, driver footage data?". The novel AI models/frameworks that make up the multifaceted driver characterisation framework are:

(a) A hybrid clustering-fuzzy framework to produce stable, reliable and interpretable driving styles (described in Section 3.3.1, page 3.3.1). The framework consists of a clustering stage that employs an ensemble of clustering and supervised learning methods on unlabelled telematics data to identify core and stable driving styles. Subsequently, the framework employs a data-driven fuzzy logic system (FLS) on the labelled data obtained from the first stage to capture, model and represent imprecision in the description of driving features and driving styles. It also extracts human-understandable explanations of the relationships between driving features and predicted driving styles. The outputs of the framework are the likelihoods of driving patterns belonging to different driving styles, such as 'calm', 'normal', 'speedy', and 'aggressive' driving styles.

(b) A hybrid deep learning framework to accurately predict driver distracted postures and affective states, described in Section 3.3.2 (page 3.3.2). The framework consists of Convolutional Neural Networks (CNNs) coupled with stacked bi-directional Recurrent Neural Networks (RNNs) to capture the spatio-temporal discriminative features of driver images. The CNN feature maps extracted from a sequence of images are fed to the RNN architecture to learn the temporal dependencies between the driving postures. The framework produces state-of-the-art classification accuracy of driver distraction postures using the popular AUC Distracted Driver Database, and state-of-the-art arousal accuracy using the popular RECOLA affective datasets.

   (c) A federated learning (FL) deep learning framework to address the challenges of data privacy, especially for driver footage data, described in Section 3.3.3 (page 3.3.3). A FL approach is implemented for the hybrid deep learning architecture described in the previous contribution. The approach predicts driver traits by processing footage data in local devices/machines and transferring the locally trained models to a central server/machine for updating the inference model. As a result, drivers' identity and privacy are safeguarded as the data remain locally. Results of the approach on the popular RECOLA affective datasets show promising affect recognition performance.

2. A 4-stage stakeholder-supported intelligent driver assessment (SIA) framework that captures and utilises information from stakeholders about the impact of contextual factors on road safety in its assessment of HGV driving behaviour (Chapter 4). The information is mostly not present in driver data or/and current AI driving assessment approaches, but at least can be partially obtained from stakeholders' experiences in the HGV sector. The capture and incorporation of the impact of contextual factors into the assessment of HGV driving allows for fairer, transparent and explainable decisions. Stages 1 and 2 of the framework produce questionnaires that capture stakeholders' views about the impact of important, up-to-date contextual factors. Stage 3 aggregates the responses from stakeholders using fuzzy sets based on the interval agreement approach to model variability in stakeholders' perception and views. The last stage defuzzifies the aggregated information and incorporates them into online driving assessments. This contribution partially answers the third research question of this thesis i.e., "how can a reliable driving risk assessment system that considers the real-world characteristics of the driving environment be developed, taking into consideration the lack of comprehensive driving risk datasets?"

3. SIA is limited to individual effects of contextual factors on road safety, which could potentially lead to highly correlated factors and overestimated assessments. In addition, it does not capture uncertainties in the characteristics of HGV driving assessment, such as, imprecise information about driving behaviours and external factors, and variability in experts' subjective views about the complex relationships between factors. To solve the aforementioned limitations of SIA, SIA is extended using a hierarchical Mamdani fuzzy inference system (Chapter 5). The new extended framework is called stakeholder-supported intelligent fuzzy driver assessment (SIFA). The framework captures information about the synergistic effects and interactions between contextual

factors from stakeholders using IF-THEN rules. Subsequently, it develops a hierarchical Mamdani fuzzy inference system to produce more realistic, interpretable and reliable decisions by fusing information about detected driver traits and perceived external conditions. The framework is evaluated using realistic HGV driving scenarios developed with the support of HGV drivers. This contribution answers the last two research questions of this thesis i.e., "how can more realistic, interpretable, fair and reliable assessments be produced that consider the real-world characteristics of HGV driving environment i.e., the synergy, interaction and uncertainties of contextual factors?" and "how can the reliability and effectiveness of the driving risk assessment system be evaluated taking into consideration the lack of multi-modal data? "

Other secondary contributions relating to road safety and driver behaviour include:

1. An extended review of the literature on intelligence-supported characterisation and assessment of HGV driving behaviours to inspire and guide future research in the area (Chapter 2).

2. A novel database of rules representing the causal relationships between HGV driving incident patterns and driving styles (Table 3.8, page 61). That is, driving rules that map the interaction between driving incidents (i.e., harsh braking, over-speeding, excessive throttling and over revving incidents) to the type of driving style e.g., 'calm', 'normal', 'speedy', and 'aggressive' driving styles.

3. A comprehensive and contemporary list of factors identified by stakeholders in the HGV industry that could potentially affect driving performance and risk (Figure 4.16, page 96). That is, factors relating to driving actions and behaviours, work life and external pressures, in-vehicle technologies, and environmental conditions.

4. A novel database of rules representing the causal relationships between contextual factors and driving risk (Figures 5.13, 5.14, 5.15 and 5.16). That is, rules that map the interaction between drivers' actions, behaviours and external conditions to the level of road safety risk e.g., 'low', 'moderate' and 'high' driving risks.

6.2 Research Findings

This section summarises the major findings of this research. The findings are grouped into computer science and road safety findings.

First, bi-directional RNNs are effective in capturing the temporal dynamics and differences between the different types of driver distraction postures and affective states using footage data. Secondly, using a federated learning implementation of machine learning models could address the issue of individual data protection, and therefore, can be applied to protect drivers' privacy and identity when processing their data. However, there is a trade-off between efficiency and privacy as non-federated strategies are still more accurate than privacy-preserving strategies. Lastly, in situations with lack of multi-modal data streams, modelling potential uncertainties in information (e.g., imprecision of information about driving behaviours and external factors, uncertainty in information produced by AI approaches, and variability in experts' subjective views about driving assessments), and the synergy between risk factors using a stakeholder-supported hierarchical fuzzy expert system produces more realistic, understandable and decomposable assessments compared to current intelligent approaches.

With regards to road safety findings, first there is always some level of driving risks regardless of the driving condition. Secondly, it was uncovered that stakeholders in the HGV sector (i.e., HGV transport managers, HGV drivers, road safety officers and researchers) generally agree on the direction of impact of contextual factors on driving performance (whether positive or negative impact), but vary in their opinions about the extent of the impact. The variations in their opinions can be attributed to differences in their roles, experiences and goals. HGV transport managers show the highest degree of disagreement with other stakeholders. The following factors were identified as the top five HGV contextual factors that impact HGV driving risk: feeling angry, aggressive driving styles, foggy weather, icy roads, and feeling tired. While attentive driving, low time pressure for delivery, calm driving, driver feeling happy and low traffic congestion are the top five factors that positively impact driving performance. The views of stakeholders also reveal that the risk associated to environmental conditions is greater than the risk associated to driving behaviours. Lastly, incorporating information about external circumstances and context (e.g., time pressure, weather conditions, road types and traffic conditions etc) into the assessment of HGV driving risk produces fairer, explainable and more realistic decisions.

6.3 Limitations and Future Work

Based on the contributions and findings of this research, we can conclude that the research questions of this thesis have been answered and the aim of this thesis has been achieved. Our resulting online driving assessment framework can automatically and accurately detect driver traits from multi-modal driver data, and assess their synergistic effects together with external circumstances on HGV driving performance and risk. Our framework, however, presents a few limitations with regards to: (1) data; and (2) methodology. This section discusses the limitations, and present potential directions to tackle the limitations and improve our framework.

The multi-modal data used in this research to evaluate the multifaceted driver behaviour characterisation framework are not entirely obtained from HGV drivers, they are not simultaneously collected and they are not related. That is, the telematics data are not affiliated to driver footage, and driver posture images are not affiliated to the facial images. This limits the development and evaluation of the framework using multi-modal data. As a result, multi-modal data that consist of telematics data, driver posture images and facial images need to be concurrently captured from drivers to provide a more realistic implementation and evaluation of the framework. To develop and evaluate an end-to-end context-aware driving risk assessment system, multi-modal data that capture driving behaviours and external conditions needs to be utilised.

With regards to the proposed intelligent context-aware HGV driving risk assessment framework, the multifaceted driver characterisation framework is limited to the classification of driver traits based on the training data. Therefore, it cannot detect new driver traits that arise due to changes in behaviour over time and changes in driving dynamics. For future work, intelligent methods that capture changes in data distribution will be examined, such as, concept drift [353] and anomaly detection techniques [354]. In addition, the state-of-the-art accuracy of federated learning to address the issue of driver privacy still requires further improvement for a trustworthy privacy-preserving solution. Non-federated processing of the raw data shows better accuracy compared to federated learning due to limited individual data at the local machines. To improve performance in the future, synthetic data generation strategies could be explored to generate more diverse data at the local machines e.g., by using generative models [355]. In addition, multi-view data sources, such as facial expressions, voice patterns, eye movements could be analysed in the local machines to provide more information about driver traits. Lastly, the framework could be integrated into real-world multi-view driver monitoring systems for further evaluation and improvement.

References

[1] Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2):179–211, 1991.

[2] Victoria Bell, Divish Rengasamy, Benjamin Rothwell, and Grazziela P Figueredo. Anomaly detection for unmanned aerial vehicle sensor data using a stacked recurrent autoencoder method with dynamic thresholding. *arXiv preprint arXiv:2203.04734*, 2022.

[3] Wayne Wu, Chen Qian, Shuo Yang, Quan Wang, Yici Cai, and Qiang Zhou. Look at boundary: A boundary-aware face alignment algorithm. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2129–2138, 2018.

[4] Department of Transport: Road Freight Statistics,United Kingdom: 2018 - 2019 (https://www.gov.uk/government/statistical-data-sets/rfs01-goods-lifted-and-distance-hauled).

[5] Global status report on road safety: World Health Organization; 2018. https://www.who.int/violence_injury_prevention/road_safety_status/2018/.

[6] Geoffrey Underwood, Peter Chapman, Sharon Wright, and David Crundall. Anger while driving. *Transportation Research Part F: traffic psychology and behaviour*, 2(1):55–68, 1999.

[7] Hannah J Foy and Peter Chapman. Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. *Applied ergonomics*, 73:90–99, 2018.

[8] WHO. Global Status Report on Road Safety: World Health Organization; 2018.

[9] Tatsuaki Osafune, Toshimitsu Takahashi, Noboru Kiyama, Tsuneo Sobue, Hirozumi Yamaguchi, and Teruo Higashino. Analysis of accident risks from driving behaviors. *International journal of intelligent transportation systems research*, 15(3):192–202, 2017.

[10] Sheila G Klauer, Thomas A Dingus, Vicki L Neale, Jeremy D Sudweeks, and David J Ramsey. Comparing real-world behaviors of drivers with high versus low rates of crashes and near crashes. Technical report, 2009.

[11] Hsin-Li Chang and Chun-Chih Yeh. Factors affecting the safety performance of bus companies—the experience of taiwan bus deregulation. *Safety Science*, 43(5-6):323–344, 2005.

[12] Chien-Ming Tseng, Ming-Shan Yeh, Li-Yung Tseng, Hsin-Hsien Liu, and Min-Chi Lee. A comprehensive analysis of factors leading to speeding offenses among large-truck drivers. *Transportation research part F: traffic psychology and behaviour*, 38:171–181, 2016.

[13] Grazziela P Figueredo, Philip R Quinlan, Mohammad Mesgarpour, Jonathan M Garibaldi, and Robert I John. A data analysis framework to rank hgv drivers. In *2015 IEEE 18th international conference on intelligent transportation systems*, pages 2001–2006. IEEE, 2015.

[14] Grazziela P. Figueredo, Utkarsh Agrawal, Jimiama M. M. Mase, Mohammad Mesgarpour, Christian Wagner, Daniele Soria, Jonathan M. Garibaldi, Peer-Olaf Siebers, and Robert I. John. Identifying Heavy Goods Vehicle Driving Styles in the United Kingdom. *IEEE Transactions on Intelligent Transportation Systems*, 2018.

[15] Amir Mehdizadeh, Mohammad Ali Alamdar Yazdi, Miao Cai, Qiong Hu, Alexander Vinel, Steven E Rigdon, Karen Davis, and Fadel M Megahed. Predicting unsafe driving risk among commercial truck drivers using machine learning: lessons learned from the surveillance of 20 million driving miles. *Accident Analysis & Prevention*, 159:106285, 2021.

[16] Hongyu Hu, Qi Wang, Ming Cheng, and Zhenhai Gao. Cost-sensitive semi-supervised deep learning to assess driving risk by application of naturalistic vehicle trajectories. *Expert Systems with Applications*, 178:115041, 2021.

[17] Antoine Hébert, Ian Marineau, Gilles Gervais, Tristan Glatard, and Brigitte Jaumard. Can we estimate truck accident risk from telemetric data using machine learning? In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1827–1836. IEEE, 2021.

[18] Lotfi A Zadeh. Fuzzy sets. *Information and control*, 8(3):338–353, 1965.

[19] Christian Wagner, Michael Smith, Ken Wallace, and Amir Pourabdollah. Generating uncertain fuzzy logic rules from surveys: capturing subjective relationships between variables from human experts. In *2015 IEEE International Conference on Systems, Man, and Cybernetics*, pages 2033–2038. IEEE, 2015.

[20] Alex B McBratney and Adrian W Moore. Application of fuzzy sets to climatic classification. *Agricultural and forest meteorology*, 35(1-4):165–185, 1985.

[21] Ranjit Biswas. An application of fuzzy sets in students' evaluation. *Fuzzy sets and systems*, 74(2):187–194, 1995.

[22] Abbas Mardani, Robert E Hooker, Seckin Ozkul, Sun Yifan, Mehrbakhsh Nilashi, Hamed Zamani Sabzi, and Goh Chin Fei. Application of decision

making and fuzzy sets theory to evaluate the healthcare and medical problems: a review of three decades of research with recent developments. *Expert Systems with Applications*, 137:202–231, 2019.

[23] Friedrich Steimann. On the use and usefulness of fuzzy sets in medical ai. *Artificial intelligence in medicine*, 21(1-3):131–137, 2001.

[24] Tien-Fu Liang. Application of fuzzy sets to manufacturing/distribution planning decisions in supply chains. *Information Sciences*, 181(4):842–854, 2011.

[25] EG Dunn, JM Keller, LA Marks, JE Ikerd, PD Gader, and LD Godsey. Extending the application of fuzzy sets to the problem of agricultural sustainability. In *Proceedings of 3rd International Symposium on Uncertainty Modeling and Analysis and Annual Conference of the North American Fuzzy Information Processing Society*, pages 497–502. IEEE, 1995.

[26] Sarah Redshaw. Theories of driver behaviour and driving emotions. In *Proceedings of the Australasian road safety research, policing and education conference*, volume 8. Monash University, 2004.

[27] Damian R Poulter, Peter Chapman, Peter A Bibby, David D Clarke, and David Crundall. An application of the theory of planned behaviour to truck driving behaviour and compliance with regulations. *Accident analysis & prevention*, 40(6):2058–2064, 2008.

[28] Heikki Summala. Traffic psychology theories: Towards understanding driving behaviour and safety factors. In *International Conference of Traffic and Transport Psychology*, 2005.

[29] Ray Fuller. Towards a general theory of driver behaviour. *Accident analysis & prevention*, 37(3):461–472, 2005.

[30] R Fuller, Catherine McHugh, and S Pender. Task difficulty and risk in the determination of driver behaviour. *European Review of Applied Psychology*, 58(1):13–21, 2008.

[31] Maya L Uhl. *The Effects of Driver Distraction on Simulated Driving: Are Multiple Resource Theory Predictions Mediated by Central Executive Involvement?* PhD thesis, California State University, Northridge, 2003.

[32] Yu Zhang, David B Kaber, Meghan Rogers, Yulan Liang, and Shruti Gangakhedkar. The effects of visual and cognitive distractions on operational and tactical driving behaviors. *Human factors*, 56(3):592–604, 2014.

[33] Robert Broström and Staffan Davidsson. Towards a model to interpret driver behaviour in terms of mismatch between real world complexity and invested effort. *Work*, 41(Supplement 1):5068–5074, 2012.

[34] M Alonso, MH Vega, and O Martín. Driving simulation study for the analysis of distraction effects in longitudinal driving behaviour. *Cognition, Technology & Work*, 14(4):283–297, 2012.

[35] Henriette Wallén Warner and Lars Åberg. Drivers' decision to speed: A study inspired by the theory of planned behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, 9(6):427–433, 2006.

[36] Kristine Malnaca. Risk homeostasis theory in traffic safety. In *Proceedings of the 21th ICTCT Workshop*. Citeseer, 2008.

[37] Leon S Robertson and Ivan Barry Pless. Does risk homoeostasis theory have implications for road safety. against. *BMJ (Clinical research ed.)*, 324(7346):1149–1152, 2002.

[38] Hamish A Deery and Brian N Fildes. Young novice driver subtypes: Relationship to high-risk behavior, traffic accident record, and simulator driving performance. *Human factors*, 41(4):628–643, 1999.

[39] C Wickens. Multiple resources and performance prediction. *Ergonomics: Psychological mechanisms and models in ergonomics*, 83, 2005.

[40] Toru Hagiwara, Takumi Kamada, Yutaro Suzuki, Toshiro Hashimoto, Tatsuya Iwasa, and Toshitake Kawai. Effect of multiple resource competition on driving performance using an in-vehicle information system. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 58, pages 1395–1399. SAGE Publications Sage CA: Los Angeles, CA, 2014.

[41] Barry H Kantowitz. Attention and mental workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 44, pages 3–456. SAGE Publications Sage CA: Los Angeles, CA, 2000.

[42] Mohammed S Majdi, Sundaresh Ram, Jonathan T Gill, and Jeffrey J Rodríguez. Drive-net: Convolutional network for driver distraction detection. In *2018 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*, pages 1–4. IEEE, 2018.

[43] Chao Yan, Frans Coenen, and Bailing Zhang. Driving posture recognition by convolutional neural networks. *IET Computer Vision*, 10(2):103–114, 2016.

[44] Whui Kim, Hyun-Kyun Choi, Byung-Tae Jang, and Jinsu Lim. Driver distraction detection using single convolutional neural network. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1203–1205. IEEE, 2017.

[45] Céline Craye and Fakhri Karray. Driver distraction detection and recognition using rgb-d sensor. *arXiv preprint arXiv:1502.00250*, 2015.

[46] Rafael A Berri, Alexandre G Silva, Rafael S Parpinelli, Elaine Girardi, and Rangel Arthur. A pattern recognition system for detecting use of mobile phones while driving. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 411–418. IEEE, 2014.

[47] Ivan D Brown. Driver fatigue. *Human factors*, 36(2):298–314, 1994.

[48] Lawrence C Barr, CY David Yang, Richard J Hanowski, and Rebecca Olson. Assessment of driver fatigue, distraction, and performance in a naturalistic setting. *Transportation research record*, 1937(1):51–60, 2005.

[49] Hal S Stern, Daniel Blower, Michael L Cohen, Charles A Czeisler, David F Dinges, Joel B Greenhouse, Feng Guo, Richard J Hanowski, Natalie P Hartenbaum, Gerald P Krueger, et al. Data and methods for studying commercial motor vehicle driver fatigue, highway safety and long-term driver health. *Accident Analysis & Prevention*, 126:37–42, 2019.

[50] Kirti Mahajan, Nagendra R Velaga, Akhilesh Kumar, and Pushpa Choudhary. Effects of driver sleepiness and fatigue on violations among truck drivers in india. *International journal of injury control and safety promotion*, 26(4):412–422, 2019.

[51] Changkun Chen and Jun Zhang. Exploring background risk factors for fatigue crashes involving truck drivers on regional roadway networks: a case control study in jiangxi and shaanxi, china. *SpringerPlus*, 5(1):582, 2016.

[52] Lee Di Milia, Naomi L Rogers, and Torbjörn Åkerstedt. Sleepiness, long distance commuting and night work as predictors of driving performance. *PLoS One*, 7(9):e45856, 2012.

[53] Pierre Philip, Jacques Taillard, E Klein, Patricia Sagaspe, Anthony Charles, WL Davies, C Guilleminault, and Bernard Bioulac. Effect of fatigue on performance measured by a driving simulator in automobile drivers. *Journal of psychosomatic research*, 55(3):197–200, 2003.

[54] Daniela Cardone, David Perpetuini, Chiara Filippini, Edoardo Spadolini, Lorenza Mancini, Antonio Maria Chiarelli, and Arcangelo Merla. Driver stress state evaluation by means of thermal imaging: A supervised machine learning approach based on ecg signal. *Applied Sciences*, 10(16):5673, 2020.

[55] Mohammad Naim Rastgoo, Bahareh Nakisa, Frederic Maire, Andry Rakotonirainy, and Vinod Chandran. Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications*, 138:112793, 2019.

[56] David Bethge, Thomas Kosch, Tobias Grosse-Puppendahl, Lewis L Chuang, Mohamed Kari, Alexander Jagaciak, and Albrecht Schmidt. Vemotion: Using driving context for indirect emotion prediction in real-time. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 638–651, 2021.

[57] Z Constantinescu, C Marinoiu, M Vladoiu, Cristian Marinoiu, and Monica Vladoiu. Driving Style Analysis Using Data Mining Techniques. Technical Report 5, 2010.

[58] Rebecca Lawton, Dianne Parker, Antony SR Manstead, and Stephen G Stradling. The role of affect in predicting social behaviors: The case of road traffic violations. *Journal of applied social psychology*, 27(14):1258–1276, 1997.

[59] Suhaila Abdul Hanan, Mark J King, and Ioni M Lewis. Are you a mindful driver? a review of the potential explanatory value of mindfulness in predicting speeding behaviour. In *Proceedings of 2010 Australasian Road Safety Research, Policing and Education Conference*. The Road Safety Research, Policing and Education Conference, 2010.

[60] Henriette Wallén Warner. *Factors influencing drivers' speeding behaviour*. PhD thesis, Acta Universitatis Upsaliensis, 2006.

[61] Peter J Cooper. The relationship between speeding behaviour (as measured by violation convictions) and crash involvement. *Journal of Safety Research*, 28(2):83–95, 1997.

[62] Michael A Regan, John D Lee, and Kristie Young. *Driver distraction: Theory, effects, and mitigation*. CRC Press, 2008.

[63] Fredrick M Streff. Driver distraction, aggression, and fatigue: a synthesis of the literature and guidelines for michigan planning. Technical report, 2000.

[64] Nadja Schömig and Barbara Metz. Three levels of situation awareness in driving with secondary tasks. *Safety Science*, 56:44–51, 2013.

[65] Kang Jiang, Feiyang Ling, Zhongxiang Feng, Kun Wang, and Cheng Shao. Why do drivers continue driving while fatigued? an application of the theory of planned behaviour. *Transportation Research Part A: Policy and Practice*, 98:141–149, 2017.

[66] Jennifer F. May. Chapter 21 - driver fatigue. In Bryan E. Porter, editor, *Handbook of Traffic Psychology*, pages 287–297. Academic Press, San Diego, 2011.

[67] Kazimierz Jamroz and Leszek Smolarek. Driver fatigue and road safety on poland's national roads. *International journal of occupational safety and ergonomics*, 19(2):297–309, 2013.

[68] Sally Maynard, Ashleigh Filtness, Karl Miller, and Fran Pilkington-Cheney. Bus driver fatigue: A qualitative study of drivers in london. *Applied Ergonomics*, 92:103309, 2021.

[69] Ran Breuer and Ron Kimmel. A deep learning perspective on the origin of facial expressions. *arXiv preprint arXiv:1705.01842*, 2017.

[70] Scott Lash. *Critique of information*. Sage, 2002.

[71] Karel A Brookhuis, Dick De Waard, and Wiel H Janssen. Behavioural impacts of advanced driver assistance systems–an overview. *European Journal of Transport and Infrastructure Research*, 1(3), 2019.

[72] HR Sneha, Mohammed Rafi, MV Manoj Kumar, Likewin Thomas, and B Annappa. Smartphone based emotion recognition and classification. In *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pages 1–7. IEEE, 2017.

[73] Sergio A Useche, Viviola Gómez Ortiz, and Boris E Cendales. Stress-related psychosocial factors at work, fatigue, and risky driving behavior in bus rapid transport (brt) drivers. *Accident Analysis & Prevention*, 104:106–114, 2017.

[74] Bin Zou, Zhenming Xiao, and Mutian Liu. Driving behavior recognition based on eeg data from a driver taking over experiment on a simulated autonomous vehicle. In *Journal of Physics: Conference Series*, volume 1550, page 042046. IOP Publishing, 2020.

[75] Grazziela P. Figueredo, Philip R. Quinlan, Mohammad Mesgarpour, Jonathan M. Garibaldi, and Robert I. John. A Data Analysis Framework to Rank HGV Drivers. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2015.

[76] Ke Wang and Jian John Lu. A two-layer risky driver recognition model with context awareness. *IEEE Access*, 9:138483–138495, 2021.

[77] Zoran Constantinescu, Cristian Marinoiu, and Monica Vladoiu. Driving style analysis using data mining techniques. *Int J of Comp Communications & Control*, 5(5):654–663, 2010.

[78] A B Ellison, S P & Greaves, and R Daniels. Profiling Drivers' Risky Behaviour Towards All Road Users. Technical report.

[79] Zejian Deng, Duanfeng Chu, Chaozhong Wu, Shidong Liu, Chen Sun, Teng Liu, and Dongpu Cao. A probabilistic model for driving-style-recognition-enabled driver steering behaviors. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2020.

[80] Zhongyang Chen, Jiadi Yu, Yanmin Zhu, Yingying Chen, and Minglu Li. D3: Abnormal driving behaviors detection and identification using smartphone sensors. In *2015 12th Annual IEEE International Conference on Sensing, Communication, and Networking, SECON 2015*, 2015.

[81] Chih-Wei Hsu, Chih-Chung Chang, Chih-Jen Lin, et al. A practical guide to support vector classification. 2003.

[82] Meinard Müller. Dynamic time warping. *Information retrieval for music and motion*, pages 69–84, 2007.

[83] Derick A Johnson and Mohan M Trivedi. Driving style recognition using a smartphone as a sensor platform. In *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1609–1615. IEEE, 2011.

[84] Nidhi Kalra and Divya Bansal. Analyzing driver behavior using smartphone sensors: a survey. *Int. J. Electron. Electr. Eng*, 7(7):697–702, 2014.

[85] Aaqib Saeed and Stojan Trajanovski. Personalized driver stress detection with multi-task neural networks using physiological signals. *arXiv preprint arXiv:1711.06116*, 2017.

[86] Neska El Haouij, Jean-Michel Poggi, Raja Ghozi, Sylvie Sevestre-Ghalila, and Mériem Jaïdane. Random forest-based approach for physiological functional variable selection for driver's stress level classification. *Statistical Methods & Applications*, 28(1):157–185, 2019.

[87] Enrique Garcia-Ceja, Venet Osmani, and Oscar Mayora. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. *IEEE journal of biomedical and health informatics*, 20(4):1053–1060, 2015.

[88] Andreas Haag, Silke Goronzy, Peter Schaich, and Jason Williams. Emotion recognition using bio-sensors: First steps towards an automatic system. In *Tutorial and research workshop on affective dialogue systems*, pages 36–48. Springer, 2004.

[89] Maria Egger, Matthias Ley, and Sten Hanke. Emotion recognition from physiological signal analysis: A review. *Electronic Notes in Theoretical Computer Science*, 343:35–55, 2019.

[90] Choubeila Maaoui and Alain Pruski. Emotion recognition through physiological signals for human-machine communication. *Cutting Edge Robotics*, 2010:317–332, 2010.

[91] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, 3(1):42–55, 2011.

[92] H Leng, Y Lin, and LA Zanzi. An experimental study on physiological parameters toward driver emotion recognition. In *International Conference on Ergonomics and Health Aspects of Work with Computers*, pages 237–246. Springer, 2007.

[93] Urvashi Agrawal, Shubhangi Giripunje, and Preeti Bajaj. Emotion and gesture recognition with soft computing tool for drivers assistance system in human centered transportation. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pages 4612–4616. IEEE, 2013.

[94] Hua Gao, Anil Yüce, and Jean-Philippe Thiran. Detecting emotional stress from facial expressions for driving safety. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 5961–5965. IEEE, 2014.

[95] Fatma Nasoz, Christine L Lisetti, and Athanasios V Vasilakos. Affectively intelligent and adaptive car interfaces. *Information Sciences*, 180(20):3817–3836, 2010.

[96] Zhiyi Ma, Marwa Mahmoud, Peter Robinson, Eduardo Dias, and Lee Skrypchuk. Automatic detection of a driver's complex mental states. In *International Conference on Computational Science and Its Applications*, pages 678–691. Springer, 2017.

[97] Albert C Cruz and Alex Rinaldi. Video summarization for expression analysis of motor vehicle operators. In *International Conference on Universal Access in Human-Computer Interaction*, pages 313–323. Springer, 2017.

[98] Björn Schuller, Manfred Lang, and Gerhard Rigoll. Recognition of spontaneous emotions by speech within automotive environment. In *Tagungsband Fortschritte der Akustik-DAGA# 2006, Braunschweig*, 2006.

[99] Christian Jones and Ing-Marie Jonsson. Using paralinguistic cues in speech to recognise emotions in older car drivers. In *Affect and Emotion in Human-Computer Interaction*, pages 229–240. Springer, 2008.

[100] Arun Sahayadhas, Kenneth Sundaraj, and Murugappan Murugappan. Detecting driver drowsiness based on sensors: a review. *Sensors*, 12(12):16937–16953, 2012.

[101] Guosheng Yang, Yingzi Lin, and Prabir Bhattacharya. A driver fatigue recognition model using fusion of multiple features. In *2005 IEEE international conference on systems, man and cybernetics*, volume 2, pages 1777–1784. IEEE, 2005.

[102] Li Shiwu, Wang Linhong, Yang Zhifa, Ji Bingkui, Qiao Feiyan, and Yang Zhongkai. An active driver fatigue identification technique using multiple physiological features. In *2011 International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, pages 733–737. IEEE, 2011.

[103] Yuxuan Yang, Zhongke Gao, Yanli Li, Qing Cai, Norbert Marwan, and Jürgen Kurths. A complex network-based broad learning system for detecting driver fatigue from eeg signals. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(9):5800–5808, 2019.

[104] Xiang Li, Nanlin Tan, Tianlei Wang, and Shuqiang Su. Detecting driver fatigue based on nonlinear speech processing and fuzzy svm. In *2014 12th International Conference on Signal Processing (ICSP)*, pages 510–515. IEEE, 2014.

[105] Ashrant Aryal, Ali Ghahramani, and Burcin Becerik-Gerber. Monitoring fatigue in construction workers using physiological measurements. *Automation in Construction*, 82:154–165, 2017.

[106] Sajid Ali Khan, Shariq Hussain, Sun Xiaoming, and Shunkun Yang. An effective framework for driver fatigue recognition based on intelligent facial expressions analysis. *Ieee Access*, 6:67459–67468, 2018.

[107] Yan Zhang and Caijian Hua. Driver fatigue recognition based on facial expression analysis using local binary patterns. *Optik*, 126(23):4501–4505, 2015.

[108] Fang Zhang, Jingjing Su, Lei Geng, and Zhitao Xiao. Driver fatigue detection based on eye state recognition. In *2017 International Conference on Machine Vision and Information Technology (CMVIT)*, pages 105–110. IEEE, 2017.

[109] Long Chen, Guojiang Xin, Yuling Liu, and Junwei Huang. Driver fatigue detection based on facial key points and lstm. *Security and Communication Networks*, 2021, 2021.

[110] Zuopeng Zhao, Nana Zhou, Lan Zhang, Hualin Yan, Yi Xu, and Zhongxin Zhang. Driver fatigue detection based on convolutional neural networks using em-cnn. *Computational intelligence and neuroscience*, 2020, 2020.

[111] S Suseela, A Karunya Abinisha, S Harini, and K Pradhiksha. Convolutional neural network based driver fatigue recognition by utilizing the facial features. In *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pages 1113–1122. IEEE, 2022.

[112] LI Zuojin, LI Renjie, LI Shengbo, WANG Wenjun, and CHENG Bo. Driver fatigue recognition based on approximated entropy and complexity of steering wheel angle. *Journal of Automotive Safety and Energy*, 7(03):279, 2016.

[113] Julie Paxion, Edith Galy, and Catherine Berthelon. Mental workload and driving. *Frontiers in psychology*, 5:1344, 2014.

[114] Jie Xie, Allaa R Hilal, and Dana Kulic. Driver distraction recognition based on smartphone sensor data. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 801–806. IEEE, 2018.

[115] Erin T Solovey, Marin Zec, Enrique Abdon Garcia Perez, Bryan Reimer, and Bruce Mehler. Classifying driver workload using physiological and driving performance data: two field studies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 4057–4066, 2014.

[116] Satoru Tokuda, Goro Obinata, Evan Palmer, and Alex Chaparro. Estimation of mental workload using saccadic eye movements in a free-viewing task. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4523–4529. IEEE, 2011.

[117] Sheila G Klauer, Thomas A Dingus, Vicki L Neale, Jeremy D Sudweeks, David J Ramsey, et al. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. 2006.

[118] Yusuf Artan, Orhan Bulan, Robert P Loce, and Peter Paul. Driver cell phone usage detection from hov/hot nir images. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 225–230, 2014.

[119] Zoubin Ghahramani. Unsupervised learning. In *Summer School on Machine Learning*, pages 72–112. Springer, 2003.

[120] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[121] Tadeusz Caliński and Jerzy Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27, 1974.

[122] John A Hartigan. Statistical theory in clustering. *Journal of classification*, 2(1):63–76, 1985.

[123] Slobodan Petrovic. A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters. In *Proceedings of the 11th Nordic Workshop of Secure IT Systems*, volume 2006, pages 53–64. Citeseer, 2006.

[124] Purnima Bholowalia and Arvind Kumar. Ebk-means: A clustering technique based on elbow method and k-means in wsn. *International Journal of Computer Applications*, 105(9), 2014.

[125] James C Bezdek. Cluster validity with fuzzy sets. 1973.

[126] Derek Greene, Alexey Tsymbal, Nadia Bolshakova, and Pádraig Cunningham. Ensemble clustering in medical diagnostics. In *Proceedings. 17th IEEE Symposium on Computer-Based Medical Systems*, pages 576–581. IEEE, 2004.

[127] Daniele Soria, Jonathan M Garibaldi, Federico Ambrogi, Andrew R Green, Des Powe, Emad Rakha, R Douglas Macmillan, Roger W Blamey, Graham Ball, Paulo JG Lisboa, et al. A methodology to identify consensus classes from clustering algorithms applied to immunohistochemical data from breast cancer patients. *Computers in biology and medicine*, 40(3):318–330, 2010.

[128] Rizwana Kalsoom and Zahid Halim. Clustering the driving features based on data streams. In *2013 16th International Multi Topic Conference, INMIC 2013*, 2013.

[129] German Castignani, Thierry Derrmann, Raphaël Frank, and Thomas Engel. Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. *IEEE Int Transp Sys Magazine*, 7(1):91–102, 2015.

[130] Eleni G Mantouka, Emmanouil N Barmpounakis, and Eleni I Vlahogianni. Identification of driving safety profiles from smartphone data using machine learning techniques. *Safety Science*, 119, 2019.

[131] Amin Mohammadnazar, Ramin Arvin, and Asad J Khattak. Classifying travelers' driving style using basic safety messages generated by connected vehicles: Application of unsupervised machine learning. *Transportation research part C: emerging technologies*, 122:102917, 2021.

[132] Jingqiu Guo, Yangzexi Liu, Lanfang Zhang, and Yibing Wang. Driving behaviour style study with a hybrid deep learning framework based on gps data. *Sustainability*, 10(7):2351, 2018.

[133] Soma Bandyopadhyay, Anish Datta, Shruti Sachan, and Arpan Pal. Unsupervised driving behavior analysis using representation learning and exploiting group-based training. *arXiv preprint arXiv:2205.07870*, 2022.

[134] Jie Chen, ZhongCheng Wu, and Jun Zhang. Driver identification based on hidden feature extraction by using adaptive nonnegativity-constrained autoencoder. *Applied Soft Computing*, 74:1–9, 2019.

[135] Mohammad Siami, Mohsen Naderpour, and Jie Lu. A mobile telematics pattern recognition framework for driving behavior extraction. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1459–1472, 2020.

[136] Eleonora Papadimitriou, Anastasia Argyropoulou, Dimitrios I Tselentis, and George Yannis. Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving. *Safety Science*, 119:91–97, 2019.

[137] Katja Kircher and Christer Ahlstrom. Predicting visual distraction using driving performance data. In *Annals of Advances in Automotive Medicine/Annual Scientific Conference*, volume 54, page 333. Association for the Advancement of Automotive Medicine, 2010.

[138] Abdulrahman Alreshidi and Mohib Ullah. Facial emotion recognition using hybrid features. In *Informatics*, volume 7, page 6. Multidisciplinary Digital Publishing Institute, 2020.

[139] Chao Yan. *Vision-based driver behaviour analysis.* The University of Liverpool (United Kingdom), 2016.

[140] Somchanok Tivatansakul, Michiko Ohkura, Supadchaya Puangpontip, and Tiranee Achalakul. Emotional healthcare system: Emotion detection by facial expressions using japanese database. In *2014 6th Computer science and electronic engineering conference (CEEC)*, pages 41–46. IEEE, 2014.

[141] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*, volume 398. John Wiley & Sons, 2013.

[142] Raymond Ghandour, Albert Jose Potams, Ilyes Boulkaibet, Bilel Neji, and Zaher Al Barakeh. Driver behavior classification system analysis using machine learning methods. *Applied Sciences*, 11(22):10562, 2021.

[143] Liu Yang, Xiaomeng Li, Wei Guan, and Shixiong Jiang. Assessing the relationship between driving skill, driving behavior and driving aggressiveness. *Journal of Transportation Safety & Security*, pages 1–17, 2020.

[144] Lipo Wang. *Support vector machines: theory and applications*, volume 177. Springer Science & Business Media, 2005.

[145] Yulan Liang, Michelle L Reyes, and John D Lee. Real-time detection of driver cognitive distraction using support vector machines. *IEEE transactions on intelligent transportation systems*, 8(2):340–350, 2007.

[146] Yanli Ma, Gaofeng Gu, Biqing Yin, Shouming Qi, Ke Chen, and Chingyao Chan. Support vector machines for the identification of real-time driving distraction using in-vehicle information systems. *Journal of Transportation Safety & Security*, pages 1–24, 2020.

[147] Zhijun Chen, Chaozhong Wu, Zhen Huang, Nengchao Lyu, Zhaozheng Hu, Ming Zhong, Yang Cheng, and Bin Ran. Dangerous driving behavior detection using video-extracted vehicle trajectory histograms. *Journal of Intelligent Transportation Systems*, 21(5):409–421, 2017.

[148] Bahzad Charbuty and Adnan Abdulazeez. Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01):20–28, 2021.

[149] Wen-Hui Chen, Yu-Chen Lin, and Wei-Hao Chen. Comparisons of machine learning algorithms for driving behavior recognition using in-vehicle can bus data. In *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*, pages 268–273. IEEE, 2019.

[150] Muhammad Zahid, Yangzhou Chen, Sikandar Khan, Arshad Jamal, Muhammad Ijaz, and Tufail Ahmed. Predicting risky and aggressive driving behavior among taxi drivers: do spatio-temporal attributes matter? *International journal of environmental research and public health*, 17(11):3937, 2020.

[151] MG Srinivasa and PS Pandian. Performance analysis of machine learning classifiers in estimating the driver s fatigue using physiological signals. *IJIREEICE*, 7(12):4–22, 2019.

[152] MS Wang, NT Jeong, KS Kim, SB Choi, SM Yang, SH You, JH Lee, and MW Suh. Drowsy behavior detection based on driving information. *International journal of automotive technology*, 17(1):165–173, 2016.

[153] Judea Pearl. Bayesian networks. 2011.

[154] Lixin Yan, Zhen Huang, Yishi Zhang, Liyan Zhang, Dunyao Zhu, and Bin Ran. Driving risk status prediction using bayesian networks and logistic regression. *Iet intelligent transport systems*, 11(7):431–439, 2017.

[155] Abdullah Rashwan. Automatic driver fatigue monitoring using hidden markov models and bayesian networks. Master's thesis, University of Waterloo, 2013.

[156] Wei Han, Wenshuo Wang, Xiaohan Li, and Junqiang Xi. Statistical-based approach for driving style recognition using bayesian probability with kernel density estimation. *IET Intelligent Transport Systems*, 13(1):22–30, 2019.

[157] Saif Al-Sultan, Ali H Al-Bayatti, and Hussein Zedan. Context-aware driver behavior detection system in intelligent transportation systems. *IEEE transactions on vehicular technology*, 62(9):4264–4275, 2013.

[158] Yulan Liang and John D Lee. Comparing support vector machines (svms) and bayesian networks (bns) in detecting driver cognitive distraction using eye movements. *Passive eye monitoring: Algorithms, applications and experiments*, pages 285–300, 2008.

[159] Veit Leonhardt, Timo Pech, and Gerd Wanielik. Fusion of driver behaviour analysis and situation assessment for probabilistic driving manoeuvre prediction. In *UR: BAN Human Factors in Traffic*, pages 223–244. Springer, 2018.

[160] Russell G Almond. Discussion: fuzzy logic: better science? or better engineering? *Technometrics*, 37(3):267–270, 1995.

[161] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: A convolutional neural-network approach. *IEEE transactions on neural networks*, 8(1):98–113, 1997.

[162] Steve Lawrence, C Lee Giles, and Ah Chung Tsoi. Convolutional neural networks for face recognition. In *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 217–222. IEEE, 1996.

[163] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017.

[164] Hyeon-Joong Yoo. Deep convolution neural networks in computer vision: a review. *IEIE Transactions on Smart Processing and Computing*, 4(1):35–43, 2015.

[165] Jawad Nagi, Frederick Ducatelle, Gianni A Di Caro, Dan Cireşan, Ueli Meier, Alessandro Giusti, Farrukh Nagi, Jürgen Schmidhuber, and Luca Maria Gambardella. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In *2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pages 342–347. IEEE, 2011.

[166] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[167] Akriti Jaiswal, A Krishnama Raju, and Suman Deb. Facial emotion detection using deep learning. In *2020 International Conference for Emerging Technology (INCET)*, pages 1–5. IEEE, 2020.

[168] Kun Han, Dong Yu, and Ivan Tashev. Speech emotion recognition using deep neural network and extreme learning machine. In *Interspeech 2014*, 2014.

[169] Luz Santamaria-Granados, Mario Munoz-Organero, Gustavo Ramirez-Gonzalez, Enas Abdulhay, and NJIA Arunkumar. Using deep convolutional neural network for emotion detection on a physiological signals dataset (amigos). *IEEE Access*, 7:57–67, 2018.

[170] Mojtaba Taherisadr, Priyanka Asnani, Scott Galster, and Omid Dehzangi. Ecg-based driver inattention identification during naturalistic driving using mel-frequency cepstrum 2-d transform and convolutional neural networks. *Smart Health*, 9:50–61, 2018.

[171] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[172] Hesham M Eraqi, Yehya Abouelnaga, Mohamed H Saad, and Mohamed N Moustafa. Driver distraction identification with an ensemble of convolutional neural networks. *Journal of Advanced Transportation*, 2019, 2019.

[173] José María Celaya-Padilla, Carlos Eric Galván-Tejada, Joyce Selene Anaid Lozano-Aguilar, Laura Alejandra Zanella-Calzada, Huizilopoztli Luna-García, Jorge Issac Galván-Tejada, Nadia Karina Gamboa-Rosales, Alberto Velez Rodriguez, and Hamurabi Gamboa-Rosales. "texting & driving" detection using deep convolutional neural networks. *Applied Sciences*, 9(15):2962, 2019.

[174] Abdu Gumaei, Mabrook Al-Rakhami, Mohammad Mehedi Hassan, Atif Alamri, Musaed Alhussein, Md Abdur Razzaque, and Giancarlo Fortino. A deep learning-based driver distraction identification framework over edge cloud. *Neural Computing and Applications*, pages 1–16, 2020.

[175] Arief Koesdwiady, Safaa M Bedawi, Chaojie Ou, and Fakhri Karray. End-to-end deep learning for driver distraction recognition. In *International Conference Image Analysis and Recognition*, pages 11–18. Springer, 2017.

[176] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[177] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[178] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[179] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[180] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.

[181] Zhen Zuo, Bing Shuai, Gang Wang, Xiao Liu, Xingxing Wang, Bing Wang, and Yushi Chen. Convolutional recurrent neural networks: Learning spatial dependencies for image representation. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 18–26, 2015.

[182] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[183] Alex Graves, Abdel-rahman Mohamed, and Geoffrey Hinton. Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. IEEE, 2013.

[184] Qingshan Liu, Feng Zhou, Renlong Hang, and Xiaotong Yuan. Bidirectional-convolutional lstm based spectral-spatial feature learning for hyperspectral image classification. *Remote Sensing*, 9(12):1330, 2017.

[185] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

[186] Furkan Omerustaoglu, C Okan Sakar, and Gorkem Kar. Distracted driver detection by combining in-vehicle and image data using deep learning. *Applied Soft Computing*, 96:106657, 2020.

[187] Shokoufeh Monjezi Kouchak and Ashraf Gaffar. Detecting driver behavior using stacked long short term memory network with attention layer. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[188] Shokoufeh Monjezi Kouchak and Ashraf Gaffar. Using bidirectional long short term memory with attention layer to estimate driver behavior. In *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, pages 315–320. IEEE, 2019.

[189] Rajkumar Theagarajan, Bir Bhanu, and Albert Cruz. Deepdriver: Automated system for measuring valence and arousal in car driver videos. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 2546–2551. IEEE, 2018.

[190] Guanglong Du, Zhiyao Wang, Boyu Gao, Shahid Mumtaz, Khamael M Abualnaja, and Cuifeng Du. A convolution bidirectional long short-term memory neural network for driver emotion recognition. *IEEE Transactions on Intelligent Transportation Systems*, 2020.

[191] Zhitao Xiao, Zhiqiang Hu, Lei Geng, Fang Zhang, Jun Wu, and Yuelong Li. Fatigue driving recognition network: Fatigue driving recognition via convolutional neural network and long short-term memory units. *IET Intelligent Transport Systems*, 13(9):1410–1416, 2019.

[192] Ming-Zhou Liu, Xin Xu, Jing Hu, and Qian-Nan Jiang. Real time detection of driver fatigue based on cnn-lstm. *IET Image Processing*, 2021.

[193] Panagiotis Tzirakis, George Trigeorgis, Mihalis A Nicolaou, Björn W Schuller, and Stefanos Zafeiriou. End-to-end multimodal emotion recognition using deep neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1301–1309, 2017.

[194] Jiyoung Lee, Sunok Kim, Seungryong Kiim, and Kwanghoon Sohn. Spatiotemporal attention based deep neural networks for emotion recognition. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1513–1517. IEEE, 2018.

[195] Pooya Khorrami, Tom Le Paine, Kevin Brady, Charlie Dagli, and Thomas S Huang. How deep neural networks can improve emotion recognition on video data. In *2016 IEEE international conference on image processing (ICIP)*, pages 619–623. IEEE, 2016.

[196] Juan DS Ortega, Mohammed Senoussaoui, Eric Granger, Marco Pedersoli, Patrick Cardinal, and Alessandro L Koerich. Multimodal fusion with deep neural networks for audio-video emotion recognition. *arXiv preprint arXiv:1907.03196*, 2019.

[197] Michel Valstar, Jonathan Gratch, Björn Schuller, Fabien Ringeval, Denis Lalanne, Mercedes Torres Torres, Stefan Scherer, Giota Stratou, Roddy Cowie, and Maja Pantic. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In *Proceedings of the 6th international workshop on audio/visual emotion challenge*, pages 3–10, 2016.

[198] Linlin Chao, Jianhua Tao, Minghao Yang, Ya Li, and Zhengqi Wen. Long short term memory recurrent neural network based multimodal dimensional emotion recognition. In *Proceedings of the 5th international workshop on audio/visual emotion challenge*, pages 65–72, 2015.

[199] Jing Han, Zixing Zhang, Nicholas Cummins, Fabien Ringeval, and Björn Schuller. Strength modelling for real-worldautomatic continuous affect recognition from audiovisual signals. *Image and Vision Computing*, 65:76–86, 2017.

[200] Qiang Yang, Yang Liu, Yong Cheng, Yan Kang, Tianjian Chen, and Han Yu. Federated learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 13(3):1–207, 2019.

[201] Yekta Said Can and Cem Ersoy. Privacy-preserving federated deep learning for wearable iot-based biomedical monitoring. *ACM Transactions on Internet Technology (TOIT)*, 21(1):1–17, 2021.

[202] Monica Bianchini and Franco Scarselli. On the complexity of neural network classifiers: A comparison between shallow and deep architectures. *IEEE transactions on neural networks and learning systems*, 25(8):1553–1565, 2014.

[203] Shizhao Sun, Wei Chen, Liwei Wang, Xiaoguang Liu, and Tie-Yan Liu. On the depth of deep neural networks: A theoretical view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[204] Yu Liang, Siguang Li, Chungang Yan, Maozhen Li, and Changjun Jiang. Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing*, 419:168–182, 2021.

[205] Jose Oramas, Kaili Wang, and Tinne Tuytelaars. Visual explanation by interpretation: Improving visual feedback capabilities of deep neural networks. *arXiv preprint arXiv:1712.06302*, 2017.

[206] Nguyen Hoang Phuong and Vladik Kreinovich. Fuzzy logic and its applications in medicine. *International journal of medical informatics*, 62(2-3):165–173, 2001.

[207] Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah. Accelerometer-based human fall detection using fuzzy entropy. In *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–7. IEEE, 2020.

[208] Arnold F Shapiro. An overview of insurance uses of fuzzy logic. *Computational Intelligence in Economics and Finance*, pages 25–61, 2007.

[209] David Enke, Manfred Grauer, and Nijat Mehdiyev. Stock market prediction with multiple regression, fuzzy type-2 clustering and neural networks. *Procedia Computer Science*, 6:201–206, 2011.

[210] Amir Pourabdollah, Christian Wagner, Michael Smith, and Ken Wallace. Real-world utility of non-singleton fuzzy logic systems: A case of environmental management. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–8. IEEE, 2015.

[211] Leonel Suganthi, S Iniyan, and Anand A Samuel. Applications of fuzzy logic in renewable energy systems–a review. *Renewable and sustainable energy reviews*, 48:585–607, 2015.

[212] Ondrej Linda, Milos Manic, Jim Alves-Foss, and Todd Vollmer. Towards resilient critical infrastructures: Application of type-2 fuzzy logic in embedded network security cyber sensor. In *2011 4th international symposium on resilient control systems*, pages 26–32. IEEE, 2011.

[213] Zack Ellerby, Josie McCulloch, Melanie Wilson, and Christian Wagner. Exploring how component factors and their uncertainty affect judgements of risk in cyber-security. In *International Conference on Critical Information Infrastructures Security*, pages 31–42. Springer, 2019.

[214] Elvia Martin Del Campo. *Design of fuzzy systems using knowledge extracted via neural networks.* PhD thesis, The University of Texas at El Paso, 2004.

[215] Ebrahim H Mamdani and Sedrak Assilian. An experiment in linguistic synthesis with a fuzzy logic controller. *International journal of man-machine studies*, 7(1):1–13, 1975.

[216] Shounak Roychowdhury and Witold Pedrycz. A survey of defuzzification strategies. *International Journal of intelligent systems*, 16(6):679–695, 2001.

[217] Ahmad Aljaafreh, Nabeel Alshabatat, and Munaf S Najim Al-Din. Driving style recognition using fuzzy logic. In *2012 IEEE International Conference on Vehicular Electronics and Safety (ICVES 2012)*, pages 460–463. IEEE, 2012.

[218] T Imkamon, P Saensom, P Tangamchit, and P Pongpaibool. Detection of hazardous driving behavior using fuzzy logic. In *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, volume 2, pages 657–660. IEEE, 2008.

[219] Sakchai Boonmee and Poj Tangamchit. Portable reckless driving detection system. In *2009 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, volume 1, pages 412–415. IEEE, 2009.

[220] Rui Araújo, Ângela Igreja, Ricardo De Castro, and Rui Esteves Araujo. Driving coach: A smartphone application to evaluate driving efficient patterns. In *2012 IEEE Intelligent Vehicles Symposium*, pages 1005–1010. IEEE, 2012.

[221] Luis Miguel Bergasa, Jesús Nuevo, Miguel A Sotelo, Rafael Barea, and María Elena Lopez. Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):63–77, 2006.

[222] Rita A Ribeiro, Antonio Falcao, Andre Mora, and Jose M Fonseca. Fif: A fuzzy information fusion algorithm based on multi-criteria decision making. *Knowledge-Based Systems*, 58:23–32, 2014.

[223] Lan Zhang, Henry Leung, and Keith CC Chan. Information fusion based smart home control system and its application. *IEEE Transactions on Consumer Electronics*, 54(3):1157–1165, 2008.

[224] Valentina Zaccaria, Moksadur Rahman, Ioanna Aslanidou, and Konstantinos Kyprianidis. A review of information fusion methods for gas turbine diagnostics. *Sustainability*, 11(22):6202, 2019.

[225] GVS Raju, Jun Zhou, and Roger A Kisner. Hierarchical fuzzy control. *International journal of control*, 54(5):1201–1216, 1991.

[226] Li-Xin Wang. Analysis and design of hierarchical fuzzy systems. *IEEE Transactions on Fuzzy systems*, 7(5):617–624, 1999.

[227] Helinä Häkkänen and Heikki Summala. Fatal traffic accidents among trailer truck drivers and accident causes as viewed by other truck drivers. *Accident Analysis & Prevention*, 33(2):187–196, 2001.

[228] Mark JM Sullman, Michelle L Meadows, and Karl B Pajo. Aberrant driving behaviours amongst new zealand truck drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(3):217–232, 2002.

[229] Bahar Öz, Türker Özkan, and Timo Lajunen. Professional and non-professional drivers' stress reactions and risky driving. *Transportation research part F: traffic psychology and behaviour*, 13(1):32–40, 2010.

[230] Hafiz Mohkum Hammad, Muhammad Ashraf, Farhat Abbas, Hafiz Faiq Bakhat, Saeed A Qaisrani, Muhammad Mubeen, Shah Fahad, and Muhammad Awais. Environmental factors affecting the frequency of road traffic accidents: a case study of sub-urban area of pakistan. *Environmental Science and Pollution Research*, 26:11674–11685, 2019.

[231] Paul M Salmon, Gemma JM Read, Vanessa Beanland, Jason Thompson, Ashleigh J Filtness, Adam Hulme, Rod McClure, and Ian Johnston. Bad behaviour or societal failure? perceptions of the factors contributing to drivers' engagement in the fatal five driving behaviours. *Applied ergonomics*, 74:162–171, 2019.

[232] Zack Ellerby, Christian Wagner, and Stephen B Broomell. Capturing richer information: On establishing the validity of an interval-valued survey response mode. *Behavior Research Methods*, pages 1–23, 2021.

[233] Wei Yan, Wang Xiang, SC Wong, Xuedong Yan, YC Li, and Wei Hao. Effects of hands-free cellular phone conversational cognitive tasks on driving stability based on driving simulation experiment. *Transportation research part F: traffic psychology and behaviour*, 58:264–281, 2018.

[234] Robert J Nowosielski, Lana M Trick, and Ryan Toxopeus. Good distractions: Testing the effects of listening to an audiobook on driving performance in simple and complex road environments. *Accident Analysis & Prevention*, 111:202–209, 2018.

[235] Samer H Hamdar, Lingqiao Qin, and Alireza Talebpour. Weather and road geometry impact on longitudinal driving behavior: Exploratory analysis using an empirically supported acceleration modeling framework. *Transportation research part C: emerging technologies*, 67:193–213, 2016.

[236] Guofa Li, Weijian Lai, Xiaoxuan Sui, Xiaohang Li, Xingda Qu, Tingru Zhang, and Yuezhi Li. Influence of traffic congestion on driver behavior in post-congestion driving. *Accident Analysis & Prevention*, 141:105508, 2020.

[237] Hongji Du, Xiaohua Zhao, Xingjian Zhang, Yunlong Zhang, and Jian Rong. Effects of fatigue on driving performance under different roadway geometries: a simulator study. *Traffic injury prevention*, 16(5):468–473, 2015.

[238] Yajie Zou, Yue Zhang, and Kai Cheng. Exploring the impact of climate and extreme weather on fatal traffic accidents. *Sustainability*, 13(1):390, 2021.

[239] Qiangqiang Shangguan, Ting Fu, Junhua Wang, Tianyang Luo, et al. An integrated methodology for real-time driving risk status prediction using naturalistic driving data. *Accident Analysis & Prevention*, 156:106122, 2021.

[240] Xiupeng Shi, Yiik Diew Wong, Michael Zhi-Feng Li, Chandrasekar Palanisamy, and Chen Chai. A feature learning approach based on xgboost for driving assessment and risk prediction. *Accident Analysis & Prevention*, 129:170–179, 2019.

[241] G. P. Figueredo, U. Agrawal, J. M. Mase, M. Mesgarpour, C. Wagner, D. Soria, J. M. Garibaldi, P.O. Siebers, and R. I. John. Identifying heavy goods vehicle driving styles in the united kingdom. *IEEE Transactions on Intelligent Transportation Systems (in press)*, tbc:tbc, 2018.

[242] Chen Wang, Lin Liu, Chengcheng Xu, and Weitao Lv. Predicting future driving risk of crash-involved drivers based on a systematic machine learning framework. *International journal of environmental research and public health*, 16(3):334, 2019.

[243] Muhammad Zahid, Yangzhou Chen, Arshad Jamal, Khalaf A Al-Ofi, and Hassan M Al-Ahmadi. Adopting machine learning and spatial analysis techniques for driver risk assessment: Insights from a case study. *International journal of environmental research and public health*, 17(14):5193, 2020.

[244] Alberto Fernández, Rubén Usamentiaga, Juan Carús, and Rubén Casado. Driver distraction using visual-based sensors and algorithms. *Sensors*, 16(11):1805, 2016.

[245] Adrian B Ellison, SP Greaves, and Rhonda Daniels. Profiling drivers' risky behaviour towards all road users. In *A safe system: expanding the reach: Australasian College of Road Safety national conference*, 2012.

[246] US DOT. Next generation simulation (ngsim) vehicle trajectories and supporting data. 2016.

[247] Yiyang Bian, Chang Heon Lee, J Leon Zhao, and Yibo Wan. A deep learning based model for driving risk assessment. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[248] Yiping Wu, Xiaohua Zhao, Jian Rong, and Yunlong Zhang. The effectiveness of eco-driving training for male professional and non-professional drivers. *Transportation research part D: transport and environment*, 59:121–133, 2018.

[249] Tiantian Chen, NN Sze, Sharon Newnam, and Lu Bai. Effectiveness of the compensatory strategy adopted by older drivers: difference between professional and non-professional drivers. *Transportation research part F: traffic psychology and behaviour*, 77:168–180, 2021.

[250] JK Caird and TJ Kline. The relationships between organizational and individual variables to on-the-job driver accidents and accident-free kilometres. *Ergonomics*, 47(15):1598–1613, 2004.

[251] Laura Pozueco, Abel Rionda, Alejandro García Pañeda, José Antonio Sánchez, Xabiel García Pañeda, Roberto García, David Melendi, and Alejandro García Tuero. Impact of on-board tutoring systems to improve driving efficiency of non-professional drivers. *IET Intelligent Transport Systems*, 11(4):196–202, 2017.

[252] Zhongxiang Feng, Miaomiao Yang, Changxi Ma, Kang Jiang, Yewei Lei, Wenjuan Huang, Zhipeng Huang, Jingjing Zhan, and Muxiong Zhou. Driving anger and its relationships with type a behavior patterns and trait anger: Differences between professional and non-professional drivers. *PLoS one*, 12(12):e0189793, 2017.

[253] Marko Maslać, Boris Antić, Krsto Lipovac, Dalibor Pešić, and Nenad Milutinović. Behaviours of drivers in serbia: Non-professional versus professional drivers. *Transportation research part F: traffic psychology and behaviour*, 52:101–111, 2018.

[254] Anna Anund, Christer Ahlström, Carina Fors, and Torbjörn Åkerstedt. Are professional drivers less sleepy than non-professional drivers? *Scandinavian Journal of Work, Environment & Health*, pages 88–95, 2018.

[255] Zack Ellerby, Josie McCulloch, John Young, and Christian Wagner. Decsys–discrete and ellipse-based response capture system. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2019.

[256] Javier Navarro and Christian Wagner. Measuring inter-group agreement on zslice based general type-2 fuzzy sets. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–7. IEEE, 2019.

[257] Christian Wagner, Simon Miller, Jonathan M Garibaldi, Derek T Anderson, and Timothy C Havens. From interval-valued data to general type-2 fuzzy sets. *IEEE Transactions on Fuzzy Systems*, 23(2):248–269, 2014.

[258] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.

[259] TR Razak, Jonathan M Garibaldi, Christian Wagner, Amir Pourabdollah, and Daniele Soria. Interpretability indices for hierarchical fuzzy systems. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2017.

[260] Ehsan Toreini, Mhairi Aitken, Kovila Coopamootoo, Karen Elliott, Carlos Gonzalez Zelaya, and Aad van Moorsel. The relationship between trust in ai and trustworthy machine learning technologies. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 272–283, 2020.

[261] Jennifer Wortman Vaughan and Hanna Wallach. A human-centered agenda for intelligible machine learning. *Machines We Trust: Getting Along with Artificial Intelligence*, 2020.

[262] Microlise. Microlise (telematics, transport and fleet management solutions). *Available at: http://www.microlise.com/, Last accessed 5th April 2021.*

[263] Yehya Abouelnaga, Hesham M Eraqi, and Mohamed N Moustafa. Real-time distracted driver posture classification. *arXiv preprint arXiv:1706.09498*, 2017.

[264] Fabien Ringeval, Andreas Sonderegger, Juergen Sauer, and Denis Lalanne. Introducing the recola multimodal corpus of remote collaborative and affective interactions. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG)*, pages 1–8. IEEE, 2013.

[265] Weirong Liu, Kunyuan Deng, Xiaoyong Zhang, Yijun Cheng, Zhiyong Zheng, Fu Jiang, and Jun Peng. A semi-supervised tri-catboost method for driving style recognition. *Symmetry*, 12(3):336, 2020.

[266] Wenshuo Wang, Junqiang Xi, Alexandre Chong, and Lin Li. Driving style classification using a semisupervised support vector machine. *IEEE Transactions on Human-Machine Systems*, 47(5):650–660, 2017.

[267] Yonggang Liu, Jiming Wang, Pan Zhao, Datong Qin, and Zheng Chen. Research on classification and recognition of driving styles based on feature engineering. *IEEE Access*, 7:89245–89255, 2019.

[268] Utkarsh Agrawal, Daniele Soria, Christian Wagner, Jonathan Garibaldi, Ian O Ellis, John MS Bartlett, David Cameron, Emad A Rakha, and Andrew R Green. Combining clustering and classification ensembles: A novel pipeline to identify breast cancer profiles. *Artificial Intelligence in Medicine*, 2019.

[269] Mohammad Mahdi Bejani and Mehdi Ghatee. Convolutional neural network with adaptive regularization to classify driving styles on smartphones. *IEEE transactions on intelligent transportation systems*, 21(2):543–552, 2019.

[270] Mohammad Shahverdy, Mahmood Fathy, Reza Berangi, and Mohammad Sabokrou. Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications*, 149:113240, 2020.

[271] Youness Moukafih, Hakim Hafidi, and Mounir Ghogho. Aggressive driving detection using deep learning-based time series classification. In *2019 IEEE International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–5. IEEE, 2019.

[272] Dominik Dörr, David Grabengiesser, and Frank Gauterin. Online driving style recognition using fuzzy logic. In *17th international IEEE conference on intelligent transportation systems (ITSC)*, pages 1021–1026. IEEE, 2014.

[273] Maen Saleh, Ahmad Aljaafreh, and Nashat Albdour. Fuzzy-based recognition model for driving styles. *IJEECS) International Journal of Electrical, Electronics and Computer Systems*, 16(01), 2013.

[274] Jorge Cordero, Jose Aguilar, Kristell Aguilar, Danilo Chávez, and Eduard Puerto. Recognition of the driving style in vehicle drivers. *Sensors*, 20(9):2597, 2020.

[275] Wenshuo Wang, Junqiang Xi, and Ding Zhao. Driving style analysis using primitive driving patterns with bayesian nonparametric approaches. *IEEE Transactions on Intelligent Transportation Systems*, 20(8):2986–2998, 2018.

[276] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1891–1898, 2014.

[277] Tong Xiao, Tian Xia, Yi Yang, Chang Huang, and Xiaogang Wang. Learning from massive noisy labeled data for image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2691–2699, 2015.

[278] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[279] Xueyun Chen, Shiming Xiang, Cheng-Lin Liu, and Chun-Hong Pan. Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geoscience and remote sensing letters*, 11(10):1797–1801, 2014.

[280] Wanli Ouyang, Xiaogang Wang, Xingyu Zeng, Shi Qiu, Ping Luo, Yonglong Tian, Hongsheng Li, Shuo Yang, Zhe Wang, Chen-Change Loy, et al. Deepid-net: Deformable deep convolutional neural networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2403–2412, 2015.

[281] CH Zhao, BL Zhang, Jie He, and Jie Lian. Recognition of driving postures by contourlet transform and random forests. *IET Intelligent Transport Systems*, 6(2):161–168, 2012.

[282] Mustafa Aljasim and Rasha Kashef. E2dr: a deep learning ensemble-based driver distraction detection with recommendations model. *Sensors*, 22(5):1858, 2022.

[283] Jiyoung Lee, Sunok Kim, Seungryong Kim, and Kwanghoon Sohn. Multi-modal recurrent attention networks for facial expression recognition. *IEEE Transactions on Image Processing*, 29:6977–6991, 2020.

[284] MAH Akhand, Shuvendu Roy, Nazmul Siddique, Md Abdus Samad Kamal, and Tetsuya Shimamura. Facial emotion recognition using transfer learning in the deep cnn. *Electronics*, 10(9):1036, 2021.

[285] Rebecca L Campbell, Renee Cloutier, Teah Marie Bynion, Annamarie Nguyen, Heidemarie Blumenthal, Matthew T Feldner, and Ellen W Leen-Feldner. Greater adolescent tiredness is related to more emotional arousal during a hyperventilation task: An area under the curve approach. *Journal of Adolescence*, 90:45–52, 2021.

[286] Ulrich Schimmack and Reisenzein Rainer. Experiencing activation: energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion*, 2(4):412, 2002.

[287] Pu Sun, Yuezun Li, Honggang Qi, and Siwei Lyu. Landmarkgan: Synthesizing faces from landmarks. *arXiv preprint arXiv:2011.00269*, 2020.

[288] Jongmoo Choi, Gerard Medioni, Yuping Lin, Luciano Silva, Olga Regina, Mauricio Pamplona, and Timothy C Faltemier. 3d face reconstruction using a single or multiple views. In *2010 20th International Conference on Pattern Recognition*, pages 3959–3962. IEEE, 2010.

[289] Yingruo Fan, Jacqueline CK Lam, and Victor OK Li. Demographic effects on facial emotion expression: an interdisciplinary investigation of the facial action units of happiness. *Scientific reports*, 11(1):1–11, 2021.

[290] Mimansa Jaiswal and Emily Mower Provost. Privacy enhanced multimodal neural representations for emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7985–7993, 2020.

[291] L-X Wang and Jerry M Mendel. Generating fuzzy rules by learning from examples. *IEEE Trans. Syst., Man, Cybern.*, 22(6):1414–1427, 1992.

[292] Lotfi A Zadeh. Fuzzy logic and approximate reasoning. *Synthese*, 30(3):407–428, 1975.

[293] Divish Rengasamy, Benjamin C Rothwell, and Grazziela P Figueredo. Towards a more reliable interpretation of machine learning outputs for safety-critical systems using feature importance fusion. *Applied Sciences*, 11(24):11854, 2021.

[294] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979.

[295] LKPJ Rdusseeun and P Kaufman. Clustering by means of medoids. In *Proceedings of the statistical data analysis based on the L1 norm conference, neuchatel, switzerland*, volume 31, 1987.

[296] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

[297] Renhong Zhao and Rakesh Govind. Defuzzification of fuzzy intervals. *Fuzzy sets and systems*, 43(1):45–55, 1991.

[298] Junyoung Chung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. Gated feedback recurrent neural networks. In *International conference on machine learning*, pages 2067–2075, 2015.

[299] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

[300] I Lawrence and Kuei Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, pages 255–268, 1989.

[301] Juanjuan Wang, Mantao Xu, Hui Wang, and Jiwu Zhang. Classification of imbalanced data by using the smote algorithm and locally linear embedding. In *2006 8th international Conference on Signal Processing*, volume 3. IEEE, 2006.

[302] Ou Stella Liang and Christopher C Yang. Determining the risk of driver-at-fault events associated with common distraction types using naturalistic driving data. *Journal of safety research*, 79:45–50, 2021.

[303] Afonso Vilaca, Pedro Cunha, and André L Ferreira. Systematic literature review on driving behavior. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–8. IEEE, 2017.

[304] Zack Ellerby, Christian Wagner, and Stephen Broomell. Capturing individuals' uncertainties–on establishing the validity of an interval-valued survey response mode. *arXiv preprint arXiv:2009.08456*, 2020.

[305] Anamika Satrawala, Arka Prokash Mazumdar, and Santosh Kumar Vipparthi. Distributed adaptive recommendation & time-stamp based estimation of driver-behaviour. In *2022 IEEE Region 10 Symposium (TENSYMP)*, pages 1–6. IEEE, 2022.

[306] Jianqing Wu and Hao Xu. The influence of road familiarity on distracted driving activities and driving operation using naturalistic driving study data. *Transportation research part F: traffic psychology and behaviour*, 52:75–85, 2018.

[307] David G Kidd, Julie Tison, Neil K Chaudhary, Anne T McCartt, and Tara D Casanova-Powell. The influence of roadway situation, other contextual factors, and driver characteristics on the prevalence of driver secondary behaviors. *Transportation research part F: traffic psychology and behaviour*, 41:1–9, 2016.

[308] Heye Huang, Yang Li, Xunjia Zheng, Jianqiang Wang, Qing Xu, and Sifa Zheng. Objective and subjective analysis to quantify influence factors of driving risk. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 4310–4316. IEEE, 2019.

[309] Zack Ellerby, Oliver Miles, Josie McCulloch, and Christian Wagner. Insights from interval-valued ratings of consumer products—a decsys appraisal. In

2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pages 1–8. IEEE, 2020.

[310] Simon Miller, Christian Wagner, Jonathan M Garibaldi, and Susan Appleby. Constructing general type-2 fuzzy sets from interval-valued data. In *2012 IEEE International Conference on Fuzzy Systems*, pages 1–8. IEEE, 2012.

[311] John Dawes. Five point vs. eleven point scales: Does it make a difference to data characteristics. *Australasian Journal of Market Research*, 10(1), 2002.

[312] Pedro S Coelho and Susana P Esteves. The choice between a fivepoint and a ten-point scale in the framework of customer satisfaction measurement. *International Journal of Market Research*, 49(3):313–339, 2007.

[313] Judith Runnels. Japanese english learner self-assessments on the cefr-j's a-level can-do statements using four and five-point response scales. *The Asian Journal of Applied Linguistics (Editor Dr Derek Chan)*, 1(2):167–77, 2014.

[314] Ana M Castaño, M Asunción Lubiano, and Antonio L García-Izquierdo. Gendered beliefs in stem undergraduates: A comparative analysis of fuzzy rating versus likert scales. *Sustainability*, 12(15):6227, 2020.

[315] Olimpia I Ban, Laurențiu Droj, Delia Tușe, Gabriela Droj, and Nicoleta Bugnar. Data processing by fuzzy methods in social sciences researches. example in hospitality industry. *International Journal of Computers, Communications and Control*, 17(2), 2022.

[316] Ewa Roszkowska and Bartłomiej Jefmański. Interval-valued intuitionistic fuzzy synthetic measure (i-vifsm) based on hellwig's approach in the analysis of survey data. *Mathematics*, 9(3):201, 2021.

[317] Ming-Shin Kuo. A novel interval-valued fuzzy mcdm method for improving airlines' service quality in chinese cross-strait airlines. *Transportation Research Part E: Logistics and Transportation Review*, 47(6):1177–1193, 2011.

[318] M Prabhu, Nabaz Nawzad Abdullah, Ramyar Rzgar Ahmed, T Nambirajan, and Sanjeevi Pandiyan. Segmenting the manufacturing industries and measuring the performance: Using interval-valued triangular fuzzy topsis method. *Complex & Intelligent Systems*, 6(3):591–606, 2020.

[319] Mei Ling Hui and Lazim Abdullah. Interval-valued intuitionistic fuzzy weighted entropy in evaluation of service quality. *International Journal of Informatics and Communication Technology*, 2(1):17–24, 2013.

[320] R Srinivas and Ajit Pratap Singh. Impact assessment of industrial wastewater discharge in a river basin using interval-valued fuzzy group decision-making and spatial approach. *Environment, Development and Sustainability*, 20(5):2373–2397, 2018.

[321] Svetlin Isaev, Mohannad Jreissat, Charalampos Makatsoris, Khaled Bachour, Josie McCulloch, and Christian Wagner. Interval-valued sensory evaluation for customized beverage product formulation and continuous manufacturing. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE, 2017.

[322] AMG Cornelissen, Jan van den Berg, Wiebe J Koops, and Uzay Kaymak. Elicitation of expert knowledge for fuzzy evaluation of agricultural production systems. *Agriculture, ecosystems & environment*, 95(1):1–18, 2003.

[323] N Amira M Saffie, Khairul A Rasmani, et al. Fuzzy delphi method: Issues and challenges. In *2016 International Conference on Logistics, Informatics and Service Sciences (LISS)*, pages 1–7. IEEE, 2016.

[324] Tufan Demirel, Nihan Çetin Demirel, and Cengiz Kahraman. Fuzzy analytic hierarchy process and its application. *Fuzzy multi-criteria decision making: Theory and applications with recent developments*, pages 53–83, 2008.

[325] Xiaoyan Hu and Philippos Mordohai. Least commitment, viewpoint-based, multi-view stereo. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, pages 531–538. IEEE, 2012.

[326] Christian Wagner and Hani Hagras. Toward general type-2 fuzzy logic systems based on zslices. *IEEE Transactions on Fuzzy Systems*, 18(4):637–660, 2010.

[327] Christian Wagner and Hani Hagras. Employing zslices based general type-2 fuzzy sets to model multi level agreement. In *2011 IEEE symposium on advances in type-2 fuzzy logic systems (T2FUZZ)*, pages 50–57. IEEE, 2011.

[328] LinkedIn. LinkedIn (A professional social network). *Available at: http://www.linkedin.com/, Last accessed 5th April 2021.*

[329] Christian Wagner, Simon Miller, and Jonathan M Garibaldi. Similarity based applications for data-driven concept and word models based on type-1 and type-2 fuzzy sets. In *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–9. IEEE, 2013.

[330] Nilesh N Karnik and Jerry M Mendel. Type-2 fuzzy logic systems: type-reduction. In *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218)*, volume 2, pages 2046–2051. Ieee, 1998.

[331] Alessandro Calvi, Maria Rosaria De Blasiis, and Claudia Guattari. An empirical study of the effects of road tunnel on driving performance. *Procedia-social and behavioral sciences*, 53:1098–1108, 2012.

[332] Francesco Bella, Alessandro Calvi, and Fabrizio D'Amico. Analysis of driver speeds under night driving conditions using a driving simulator. *Journal of safety research*, 49:45–e1, 2014.

[333] Cheng-qiu Xie and Dianne Parker. A social psychological approach to driving violations in two chinese cities. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(4):293–308, 2002.

[334] JCF De Winter and D Dodou. The driver behaviour questionnaire as a predictor of accidents: A meta-analysis. *Journal of safety research*, 41(6):463–470, 2010.

[335] Fanxing Meng, Shuling Li, Lingzhi Cao, Musen Li, Qijia Peng, Chunhui Wang, and Wei Zhang. Driving fatigue in professional drivers: a survey of truck and taxi drivers. *Traffic injury prevention*, 16(5):474–483, 2015.

[336] Article by AJ Dellinger. Amazon's delivery drivers face pressure that leads to accidents and deaths, new reports highlight. *Available at: https://www.mic.com/p/amazons-delivery-drivers-face-pressure-that-leads-to-accidents-deaths-new-reports-highlight-18729406,Last accessed 25th May 2021.*

[337] Lerer Article by McGee and Associates. Danger and Delivery Drivers. *Available at: https://www.longbeachinjuryfirm.com/blog/2019/december/danger-and-delivery-drivers/,Last accessed 31st May 2021.*

[338] Pranamesh Chakraborty, Yaw Okyere Adu-Gyamfi, Subhadipto Poddar, Vesal Ahsani, Anuj Sharma, and Soumik Sarkar. Traffic congestion detection from camera images using deep convolution neural networks. *Transportation Research Record*, 2672(45):222–231, 2018.

[339] Nimish Agarwal, Aman Jangid, Ashutosh Sharma, Nikhil Kumar, Mudit Kumar, Pankaj Kumar, and Pranamesh Chakraborty. Camera-based smart traffic state detection in india using deep learning models. In *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, pages 690–696. IEEE, 2021.

[340] Stephen Stradling, Emma Collins, Chris Eynon, and Pat McLeod. Driver behaviour on rural roads in scotland. In *Behavioural Research in Road Safety 2008 Eighteenth Seminar*, page 234, 2010.

[341] Mohamed R Ibrahim, James Haworth, and Tao Cheng. Weathernet: Recognising weather and visual conditions from street-level images using deep residual learning. *ISPRS International Journal of Geo-Information*, 8(12):549, 2019.

[342] Jingming Xia, Dawei Xuan, Ling Tan, and Luping Xing. Resnet15: Weather recognition on traffic road with deep convolutional neural network. *Advances in Meteorology*, 2020, 2020.

[343] Bin Zhao, Xuelong Li, Xiaoqiang Lu, and Zhigang Wang. A cnn–rnn architecture for multi-label weather recognition. *Neurocomputing*, 322:47–57, 2018.

[344] Tajul Rosli Razak, Jonathan M Garibaldi, Christian Wagner, Amir Pourabdollah, and Daniele Soria. Toward a framework for capturing interpretability

of hierarchical fuzzy systems—a participatory design approach. *IEEE Transactions on Fuzzy Systems*, 29(5):1160–1172, 2020.

[345] Basma Soua, Amel Borgi, and Moncef Tagina. An ensemble method for fuzzy rule-based classification systems. *Knowledge and information systems*, 36(2):385–410, 2013.

[346] Tajul Rosli Razak, Shukor Sanim Mohd Fauzi, Ray Adderley JM Gining, Mohammad Hafiz Ismail, and Ruhaila Maskat. Hierarchical fuzzy systems: Interpretability and complexity. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, 9(2):478–489, 2021.

[347] Jens Rasmussen. The role of hierarchical knowledge representation in decisionmaking and system management. *IEEE Transactions on systems, man, and cybernetics*, (2):234–243, 1985.

[348] Antonio Ginart, Gustavo Sanchez, I Links, and G Back. Fast defuzzification method based on centroid estimation. *Applied Modelling and Simulation*, 58(1):20–25, 2002.

[349] Odd André Hjelkrem and Eirin Olaussen Ryeng. Chosen risk level during car-following in adverse weather conditions. *Accident Analysis & Prevention*, 95:227–235, 2016.

[350] A Abdi Kordani, Omid Rahmani, Amir Saman Abdollahzadeh Nasiri, and Sid Mohammad Boroomandrad. Effect of adverse weather conditions on vehicle braking distance of highways. *Civil Engineering Journal*, 4(1):46–57, 2018.

[351] Amanda K Emo, Gerald Matthews, and Gregory J Funke. The slow and the furious: Anger, stress and risky passing in simulated traffic congestion. *Transportation research part F: traffic psychology and behaviour*, 42:1–14, 2016.

[352] Olga Vl Bitkina, Jungyoon Kim, Jangwoon Park, Jaehyun Park, and Hyun K Kim. Identifying traffic context using driving stress: A longitudinal preliminary case study. *Sensors*, 19(9):2152, 2019.

[353] Indrė Žliobaitė, Mykola Pechenizkiy, and Joao Gama. An overview of concept drift applications. *Big data analysis: new algorithms for a new society*, pages 91–114, 2016.

[354] Mai H Hassan, Ali Tizghadam, and Alberto Leon-Garcia. Spatio-temporal anomaly detection in intelligent transportation systems. *Procedia Computer Science*, 151:852–857, 2019.

[355] Peter Eigenschink, Stefan Vamosi, Ralf Vamosi, Chang Sun, Thomas Reutterer, and Klaudius Kalcher. Deep generative models for synthetic data. *ACM Computing Surveys*, 2021.

Appendix A

Appendix

A.1 Questionnaire consisting of questions that ask participants to provide their opinions or ratings about the impact of HGV contextual factors on HGV driving performance.

Page 2: Questionnaire

Your experience

2. What is your occupation? ✱ *Required*

- ◌ Fleet manager
- ◌ Driver
- ◌ Road safety professional
- ◌ Academics
- ◌ Researcher
- ◌ Other

2.a. If you selected Other, please specify:

3. How many years of experience do you have in fleet management, road safety, transport or research? ✱ *Required*

Please enter a number.

0

4. Do you currently hold a HGV license? ✱ *Required*

- ◌ Yes
- ◌ No

5. Which of the sectors below best describes your current operation? ✱ *Required*

- ◌ Single drops
- ◌ Multi drops
- ◌ Urban deliveries
- ◌ Trucking

- ◌ N/A
- ◌ Other

5.a. If you selected Other, please specify:

Response Instructions: How to complete the questionnaire

Consider the following question: **What is the impact of a driver feeling happy on their driving performance?**

- If you are **very certain** that a driver feeling happy has a strong positive impact on their performance, then select **'9'**.

	Strong Negative Impact (1)	2	Negative Impact (3)	4	No Impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
Happy	☐	☐	☐	☐	☐	☐	☐	☐	☑

- If you think a driver feeling happy has a positive impact on their performance but you a **little uncertain** about the level of impact, you can select a range to represent your uncertainty. For example, **min='7'** and **max='9'** to represent the **range of uncertainty of the impact.**

	Strong Negative Impact (1)	2	Negative Impact (3)	4	No Impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
Happy	☐	☐	☐	☐	☐	☐	☑	☐	☑

- If you think a driver feeling happy has a positive impact overall but also think that it may have no impact as well (**more uncertainty**). You can select a wider range **min='5'** and **max='9'** to represent your uncertainty.

	Strong Negative Impact (1)	2	Negative Impact (3)	4	No Impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
Happy	☐	☐	☐	☐	☑	☐	☐	☐	☑

Questionnaire

Please if you understood the above instructions, proceed to answer the following questions.

6. **Considering the following moods of a driver, how do you think they could affect their driving performance?** ✱ *Required*

	Strong Negative Impact (1)	2	Negative Impact (3)	4	No Impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
Sad	☐	☐	☐	☐	☐	☐	☐	☐	☐
Angry	☐	☐	☐	☐	☐	☐	☐	☐	☐
Happy	☐	☐	☐	☐	☐	☐	☐	☐	☐
Tired	☐	☐	☐	☐	☐	☐	☐	☐	☐
Energetic	☐	☐	☐	☐	☐	☐	☐	☐	☐
Distracted	☐	☐	☐	☐	☐	☐	☐	☐	☐
Attentive	☐	☐	☐	☐	☐	☐	☐	☐	☐
Calm	☐	☐	☐	☐	☐	☐	☐	☐	☐
Aggressive	☐	☐	☐	☐	☐	☐	☐	☐	☐
Confident	☐	☐	☐	☐	☐	☐	☐	☐	☐
Insecure	☐	☐	☐	☐	☐	☐	☐	☐	☐

7. **Considering the following periods of delivery, how do you think they could affect a driver's performance?** ✱ *Required*

	Strong Negative Impact (1)	2	Negative Impact (3)	4	No Impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
Start of shift	☐	☐	☐	☐	☐	☐	☐	☐	☐
Mid-shift	☐	☐	☐	☐	☐	☐	☐	☐	☐
End of shift	☐	☐	☐	☐	☐	☐	☐	☐	☐

8. **Considering the following external pressures, how do you think they could affect a driver's performance?** ✱ *Required*

	Strong Negative Impact (1)	2	Negative Impact (3)	4	No Impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
High traffic	☐	☐	☐	☐	☐	☐	☐	☐	☐

Time pressure for delivery	☐	☐	☐	☐	☐	☐	☐	☐	☐
Low traffic	☐	☐	☐	☐	☐	☐	☐	☐	☐

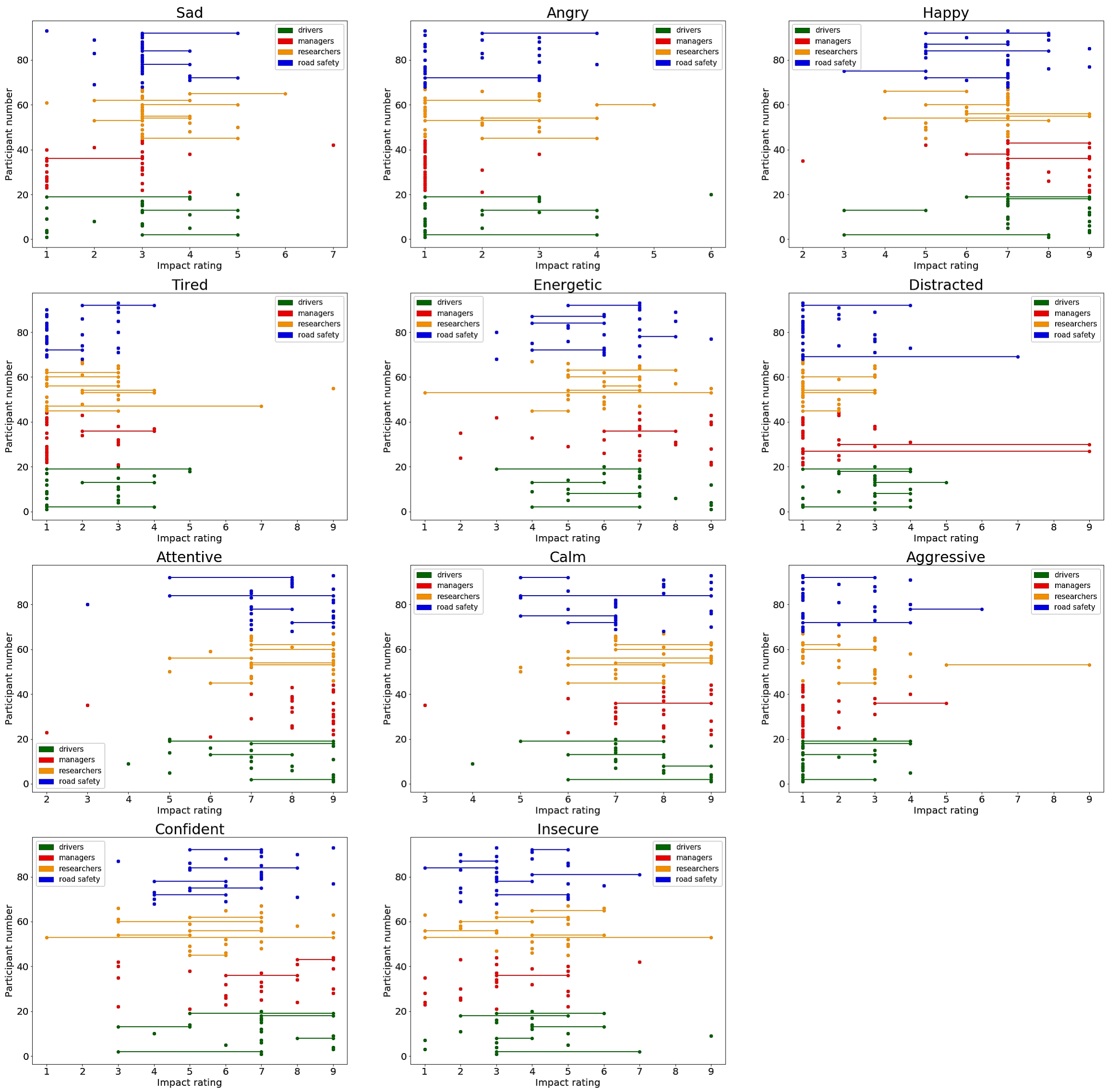
9. **Considering the following technologies, how do you think they could affect a driver's performance?**
✱ *Required*

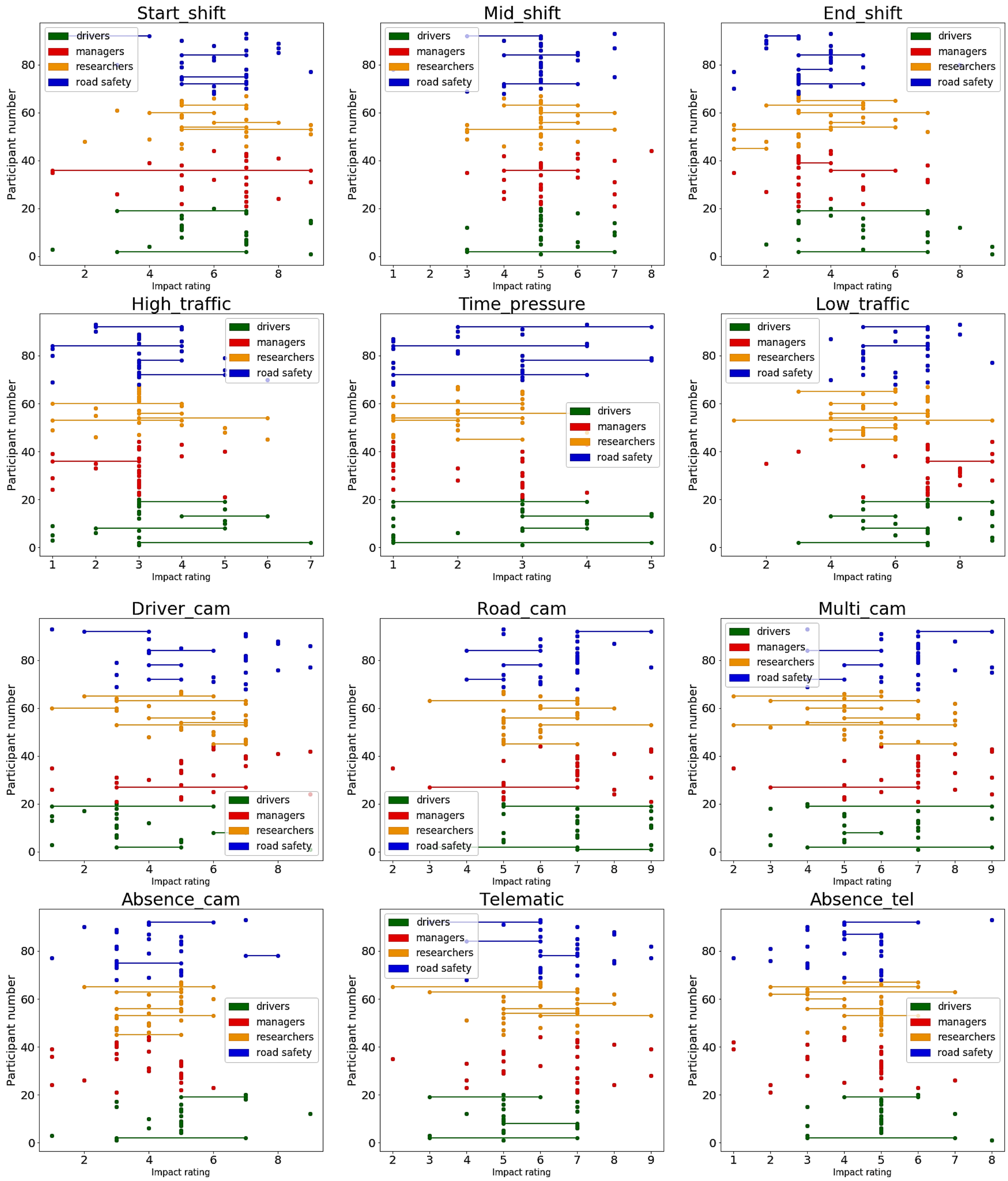
	Strong Negative Impact (1)	2	Negative impact (3)	4	No impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
Driver-facing cameras	☐	☐	☐	☐	☐	☐	☐	☐	☐
Road-facing cameras	☐	☐	☐	☐	☐	☐	☐	☐	☐
Multi-camera systems	☐	☐	☐	☐	☐	☐	☐	☐	☐
Absence of cameras	☐	☐	☐	☐	☐	☐	☐	☐	☐
Telematics	☐	☐	☐	☐	☐	☐	☐	☐	☐
Absence of telematics	☐	☐	☐	☐	☐	☐	☐	☐	☐

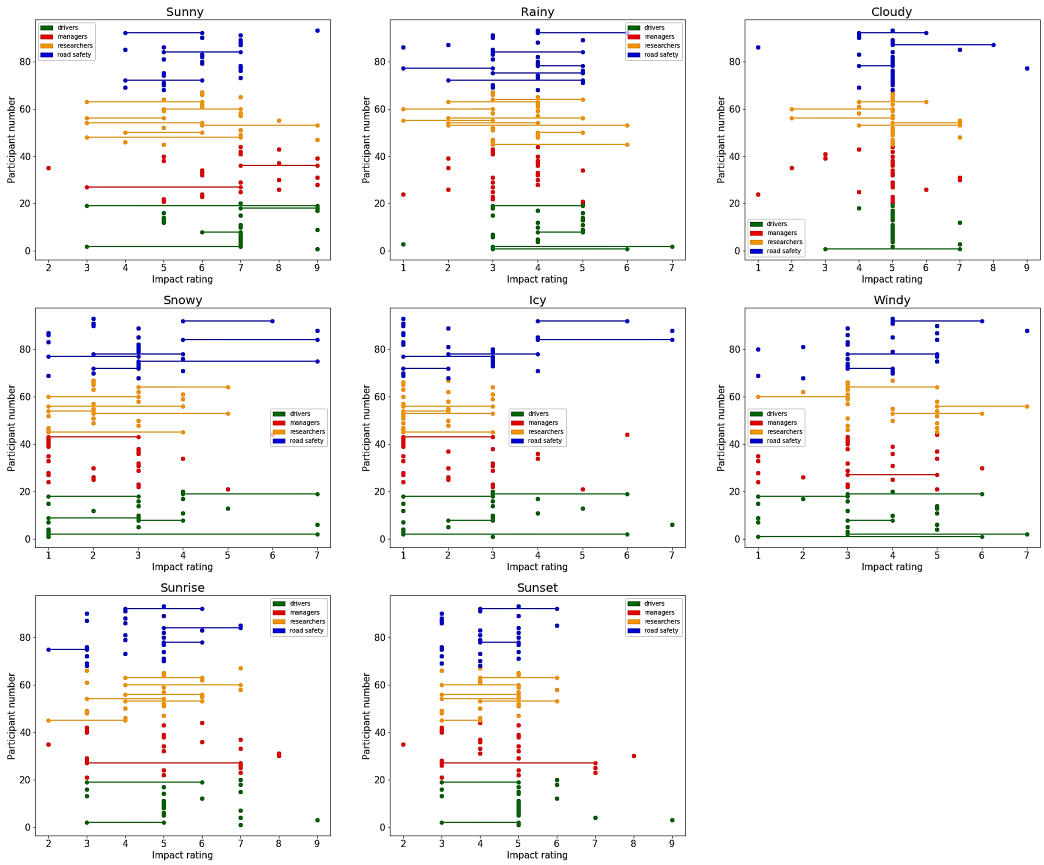
10. **Considering the following weather conditions, how do you think they could affect a driver's performance?** ✱ *Required*

	Strong Negative Impact (1)	2	Negative impact (3)	4	No impact (5)	6	Positive Impact (7)	8	Strong Positive Impact (9)
Sunny	☐	☐	☐	☐	☐	☐	☐	☐	☐
Rainy	☐	☐	☐	☐	☐	☐	☐	☐	☐
Cloudy	☐	☐	☐	☐	☐	☐	☐	☐	☐
Snowy	☐	☐	☐	☐	☐	☐	☐	☐	☐
Icy	☐	☐	☐	☐	☐	☐	☐	☐	☐
Windy	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sunrise	☐	☐	☐	☐	☐	☐	☐	☐	☐
Sunset	☐	☐	☐	☐	☐	☐	☐	☐	☐

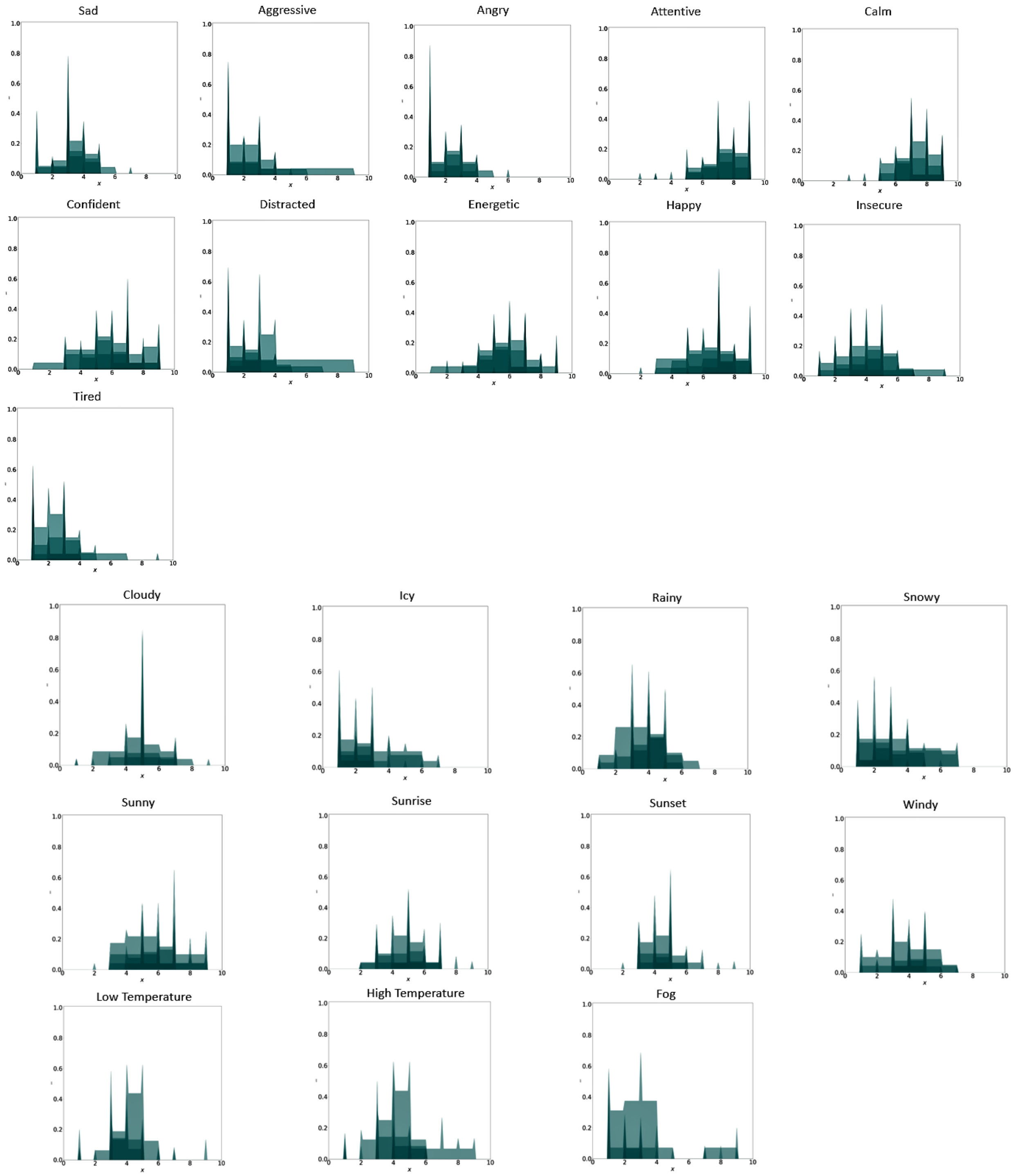
A.2 Line graphs showing individual responses from 93 experts about the impact of HGV contextual factors on HGV driving performance.

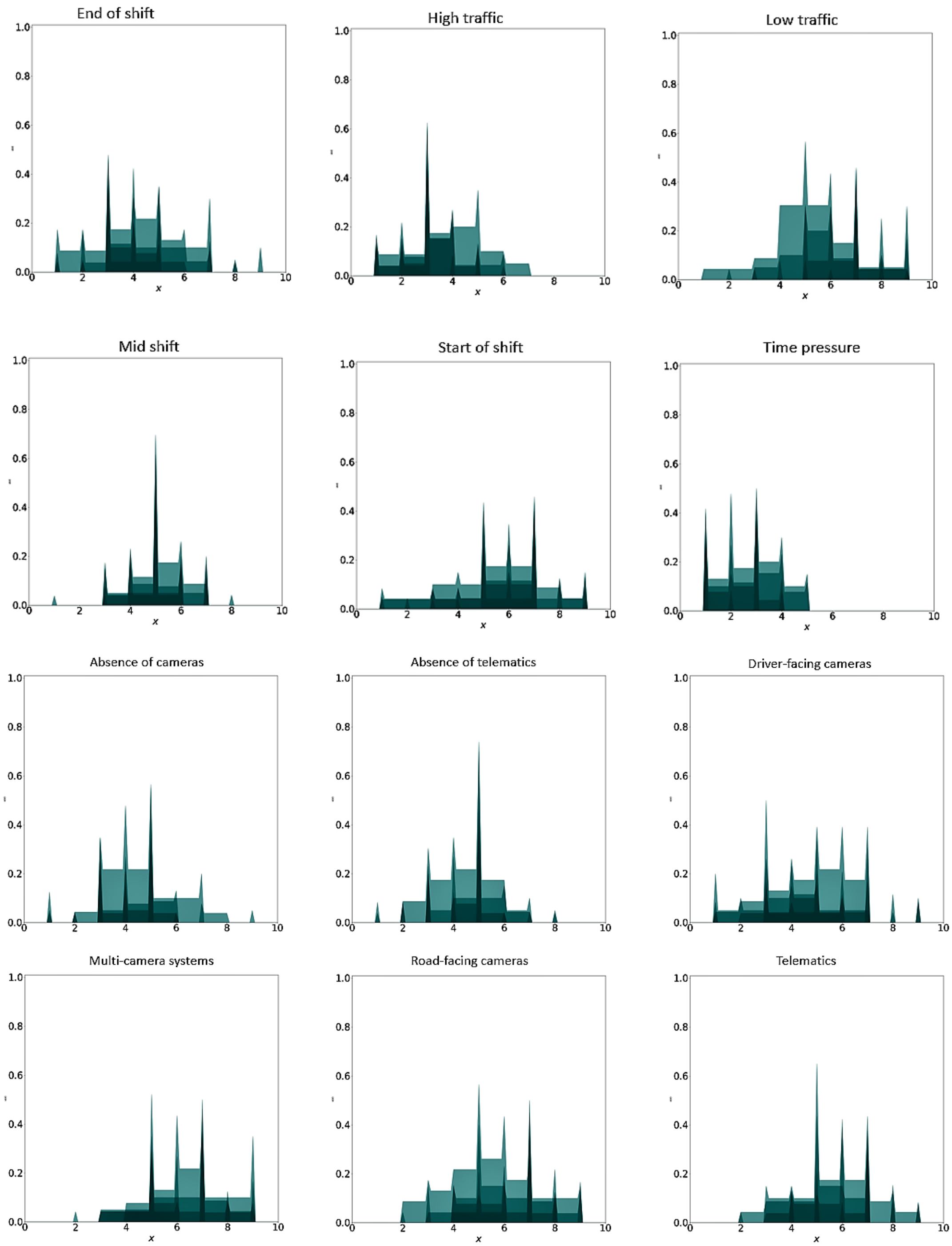



Start_shift
Mid_shift
End_shift
High_traffic
Time_pressure
Low_traffic
Driver_cam
Road_cam
Multi_cam
Absence_cam
Telematic
Absence_tel
Participant number
Impact rating
drivers
managers
researchers
road safety


Sunny
Rainy
Cloudy
Snowy
Icy
Windy
Sunrise
Sunset
drivers
managers
researchers
road safety
Participant number
Impact rating

A.3 General type-2 fuzzy sets generated from IAA fuzzy sets of HGV contextual factors.




End of shift
High traffic
Low traffic
Mid shift
Start of shift
Time pressure
Absence of cameras
Absence of telematics
Driver-facing cameras
Multi-camera systems
Road-facing cameras
Telematics

Appendix B

Appendix

B.1 Questionnaire consisting of questions that ask participants to provide their opinions or ratings about the impact of the interaction between driver traits and external factors on HGV driving risk.

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Road is **icy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☐	☑	☑

- **Participant C** thinks the level of disturbance on HGV driving caused by the conditions ranges from '**moderate**' to '**very high**'. So, selects **min='3'** and **max='5'** to represent the range of the level of external disturbance.

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Road is **icy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☑	☐	☑

Questions

Please answer the following questions by selecting a single response or a range (min-max) depending on how certain is the level of disturbance of different external conditions occuring at the same time.

4. **What is the level of external disturbance on HGV driving if the following conditions occur at the same time?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Weather is **sunny;** Traffic Congestion is **moderate;** Road is a **motorway**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **moderate;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **moderate;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **low;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **high;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **low;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **moderate;** Road is an **undivided highway**	☐	☐	☐	☐	☐

Road is **icy;** Traffic Congestion is **low;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **low;** Road is an **intersection**	☐	☐	☐	☐	☐

5. **What is the level of external disturbance on HGV driving if the following conditions occur at the same time?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Road is **icy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **high;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **high;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **low;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **low;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **high;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **high;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **low;** Road in an **urban area**	☐	☐	☐	☐	☐

The following questions require you to provide your opinion about the overall level of risk (impact on road safety) by considering risks from driving behaviours (driving risk) and external conditions (external disturbance).

We categorise **driving risk** into **'low', 'moderate'** and **'high' ,** and the **level of external disturbance** into **'low', 'moderate' and 'high'.**

Due to the complexity of assessing the overall risk, we have provided a two-point selection scale for participants to either select a single response or a a range (min-max) depending on how certain is the overall level of risk.

6. **What is the overall level of risk considering the following categories of driving risk and external disturbance?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Driving risk is **moderate;** External disturbance is **moderate**	☐	☐	☐	☐	☐

Driving risk is **high;** External disturbance is **moderate**	☐	☐	☐	☐	☐
Driving risk is **low;** External disturbance is **low**	☐	☐	☐	☐	☐
Driving risk is **high;** External disturbance is **high**	☐	☐	☐	☐	☐
Driving risk is **low;** External disturbance is **high**	☐	☐	☐	☐	☐
Driving risk is **moderate;** External disturbance is **high**	☐	☐	☐	☐	☐
Driving risk is **low;** External disturbance is **moderate**	☐	☐	☐	☐	☐
Driving risk is **moderate;** External disturbance is **low**	☐	☐	☐	☐	☐
Driving risk is **high;** External disturbance is **low**	☐	☐	☐	☐	☐

	Very low risk (1)	2	Low risk (3)	4	Moderate risk (5)	6	High risk (7)	8	Very high risk (9)
Emotion **negative;** Distraction **high;** Driving **aggressive**									
Emotion **positive;** Distraction **low;** Driving **calm**									
Emotion **positive;** Distraction **high;** Driving **aggressive**									
Emotion **negative;** Distraction **low;** Driving **aggressive**									
Emotion **positive;** Distraction **low;** Driving **aggressive**									
Emotion **negative;** Distraction **low;** Driving **normal**									
Emotion **negative;** Distraction **high;** Driving **calm**									
Emotion **positive;** Distraction **high;** Driving **normal**									
Emotion **positive;** Distraction **low;** Driving **normal**									

8. **What is the impact on road safety (risk) if the following behaviours occur at the same time?** ✱ *Required*

	Very low risk (1)	2	Low risk (3)	4	Moderate risk (5)	6	High risk (7)	8	Very high risk (9)
Emotion **negative;** Distraction **high;** Driving **normal**									

Emotion **negative;** Distraction **low;** Driving **calm**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **positive;** Distraction **high;** Driving **calm**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **negative;** Distraction **moderate;** Driving **aggressive**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **positive;** Distraction **moderate;** Driving **calm**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **neutral;** Distraction **high;** Driving **aggressive**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **neutral;** Distraction **low;** Driving **calm**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **negative;** Distraction **moderate;** Driving **calm**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **neutral;** Distraction **moderate;** Driving **normal**	☐	☐	☐	☐	☐	☐	☐	☐	☐

9. **What is the impact on road safety (risk) if the following behaviours occur at the same time?** ✱ *Required*

	Very low risk (1)	2	Low risk (3)	4	Moderate risk (5)	6	High risk (7)	8	Very high risk (9)
Emotion **neutral;** Distraction **low;** Driving **normal**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **positive;** Distraction **moderate;** Driving **normal**	☐	☐	☐	☐	☐	☐	☐	☐	☐

Emotion **negative;** Distraction **moderate;** Driving **normal**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **neutral;** Distraction **high**,**;** Driving **normal**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **neutral;** Distraction **low;** Driving **aggressive**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **neutral;** Distraction **high;** Driving **calm**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **neutral;** Distraction **moderate;** Driving **calm**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **neutral;** Distraction **moderate;** Driving **aggressive**	☐	☐	☐	☐	☐	☐	☐	☐	☐
Emotion **positive;** Distraction **moderate;** Driving **aggressive**	☐	☐	☐	☐	☐	☐	☐	☐	☐

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Road is **icy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☐	√	√

- **Participant C** thinks the level of disturbance on HGV driving caused by the conditions ranges from '**moderate**' to '**very high**'. So, selects **min='3'** and **max='5'** to represent the range of the level of disturbance.

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Road is **icy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	√	☐	√

Questions

Please answer the following questions by selecting a single response or a range (min-max) depending on how certain is the level of disturbance of different external conditions occuring at the same time.

4. **What is the level of external disturbance on HGV driving if the following conditions occur at the same time?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Weather is **rainy;** Traffic Congestion is **low;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **moderate;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **high;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **moderate;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **high;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **low;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☐	☐	☐

Weather is **foggy;** Traffic Congestion is **high;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **high;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **moderate;** Road in an **urban area**	☐	☐	☐	☐	☐

5. **What is the level of external disturbance on HGV driving if the following conditions occur at the same time?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Road is **icy;** Traffic Congestion is **low;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **moderate;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **high;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **low;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **high;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **low;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **high;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **moderate;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **moderate;** Road is a **motorway**	☐	☐	☐	☐	☐

6. **What is the level of external disturbance on HGV driving if the following conditions occur at the same time?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Road is **icy;** Traffic Congestion is **moderate;** Road is an **intersection**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **high;** Road in a **rural area**	☐	☐	☐	☐	☐

Weather is **rainy;** Traffic Congestion is **low;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **moderate;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **low;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **moderate;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **high;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **moderate;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **high;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **moderate;** Road in a **rural area**	☐	☐	☐	☐	☐

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Road is **icy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☐	√	√

- **Participant C** thinks the level of disturbance on HGV driving caused by the conditions ranges from '**moderate**' to '**very high**'. So, selects **min='3'** and **max='5'** to represent the range of the level of external disturbance.

	Very Low (1)	Low (2)	Moderate (3)	High (4)	Very High (5)
Road is **icy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	√	☐	√

Questions

Please answer the following questions by selecting a single response or a range (min-max) depending on how certain is the level of disturbance of different external conditions occuring at the same time.

4. **What is the level of external disturbance on HGV driving if the following conditions occur at the same time?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Road is **icy;** Traffic Congestion is **low;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **moderate;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **moderate;** Road is an **intersection**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **high;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **low;** Road is an **intersection**	☐	☐	☐	☐	☐

Weather is **sunny;** Traffic Congestion is **high;** Road is in **urban area**	☐	☐	☐	☐	☐

5. **What is the level of external disturbance on HGV driving if the following conditions occur at the same time?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Weather is **foggy;** Traffic Congestion is **moderate;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **low;** Road is a **motorway**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **low;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **moderate;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **moderate;** Road in a **rural area**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **low;** Road in a **urban area**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **moderate;** Road is a **motorway**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **moderate;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **high;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **high;** Road in a **rural area**	☐	☐	☐	☐	☐

6. **What is the level of external disturbance on HGV driving if the following conditions occur at the same time?** ✱ *Required*

	Very low (1)	Low (2)	Moderate (3)	High (4)	Very high (5)
Weather is **rainy;** Traffic Congestion is **high;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **moderate;** Road is an **undivided highway**	☐	☐	☐	☐	☐

Weather is **snowy;** Traffic Congestion is **moderate;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **snowy;** Traffic Congestion is **high;** Road is an **undivided highway**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **high;** Road is an **intersection**	☐	☐	☐	☐	☐
Weather is **sunny;** Traffic Congestion is **low;** Road in an **urban area**	☐	☐	☐	☐	☐
Road is **icy;** Traffic Congestion is **high;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **rainy;** Traffic Congestion is **moderate;** Road in an **urban area**	☐	☐	☐	☐	☐
Weather is **foggy;** Traffic Congestion is **high;** Road is an **undivided highway**	☐	☐	☐	☐	☐

B.2 Questionnaire asking drivers to validate and revise user stories representing real-world HGV driving scenarios.

⏱ 30 minutes

Evaluating Heavy Goods Vehicle Driving User Stories

The survey will take approximately 6 minutes to complete. Ethics Ref: CS-2020-R9

We have developed a HGV driving risk-assessment system that automatically evaluates the risks associated with driving styles, drivers' moods, weather and road conditions. We wish to evaluate the performance of our system using real-world driving scenarios.

This short survey presents user stories of driving scenarios. Please can you determine if the user stories are typical real driving scenarios or are unrealistic (not possible), and can you classify the driving styles in the different user stories. If you believe the user story is unrealistic, please kindly provide suggestions on how to improve them.

1. A HGV driver is travelling on a sunny motorway with about 100 vehicles per lane of the road. The driver has a total of 2 seconds of driving at the maximum speed limit. The driver is attentive and well rested.

   On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5

   Unrealistic Very Possible

2. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

3. How will you describe the driving style in the above user story?

- ◯ Cautious
- ◯ Normal
- ◯ Reckless
- ◯ Other

4. A HGV driver is approaching a busy roundabout in a foggy weather. The driver reduces their speed well below the maximum speed limit of the road. The driver is attentive and well rested.

   On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5
Unrealistic				Very Possible

5. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

6. How will you describe the driving style in the above user story?

- ◯ Cautious
- ◯ Normal
- ◯ Reckless
- ◯ Other

7. A HGV driver is travelling in a rainy undivided highway with 50 vehicles on the road. The driver has a total of 20 seconds of driving at the maximum speed limit. The driver is well rested but distracted.

   On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5
Unrealistic				Very Possible

8. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

9. How will you describe the driving style in the above user story?

- ◯ Cautious
- ◯ Normal
- ◯ Reckless
- ◯ Other

10. A HGV driver is approaching a clear roundabout in a sunny weather. The driver maintains their speed. The driver is not well rested and in a negative mood.

    On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5

Unrealistic Very Possible

11. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

12. How will you describe the driving style in the above user story?

- ◯ Cautious
- ◯ Normal
- ◯ Reckless
- ◯ Other

13. A HGV driver is travelling on a sunny undivided highway with about 100 vehicles per lane of the road. The driver has a total of 6 harsh braking incidents. The driver is distracted and frustrated by their in-vehicle delivery management system.

    On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5

Unrealistic Very Possible

14. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

15. How will you describe the driving style in the above user story?

- Calm
- Normal
- Aggressive
- Other

16. A HGV driver is travelling in a foggy motorway with about 30 vehicles per lane of the road. The driver produces a total of 1 harsh braking incidents. The driver is distracted and in a negative mood.

    On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5

Unrealistic Very Possible

17. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

18. How will you describe the driving style in the above user story?

- Calm
- Normal
- Aggressive
- Other

19. A HGV driver is travelling on a icy rural road with with about 20 vehicles on the road. The driver produces a total of 2 rash overtaking incidents. The driver is attentive but angry due to pressure from their manager to complete a job.

    On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5

Unrealistic Very Possible

20. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

21. How will you describe the driving style in the above user story?

- Calm
- Normal
- Aggressive
- Other

22. A HGV driver is travelling in a snow urban road with with about 50 vehicles on the road. The driver produces a total of 5 harsh acceleration incidents. The driver is attentive and well rested.

    On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5

Unrealistic Very Possible

23. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

24. How will you describe the driving style in the above user story?

- Calm
- Normal
- Aggressive
- Other

25. A HGV driver is travelling on a sunny rural road with with about 5 vehicles on the road. The driver produces a total of 1 harsh acceleration incidents. The driver is in a positive mood but distracted.

    On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5

Unrealistic Very Possible

26. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

27. How will you describe the driving style in the above user story?

- Calm
- Normal
- Aggressive
- Other

28. A HGV driver is travelling on a sunny urban road with about 10 vehicles on the road. The driver produces a total of 1 harsh acceleration incidents. The driver is attentive but frustrated by the amount of jobs assigned to them.

    On a scale of 1 to 5 (unrealistic to very possible), what will you consider the above driving scenario?

1	2	3	4	5

Unrealistic Very Possible

29. If you think the above driving scenario is unrealistic (below 3), please how can we improve the user story to make it more realistic?

30. How will you describe the driving style in the above user story?

- Calm
- Normal
- Aggressive
- Other

B.3 Questionnaire consisting of questions that ask drivers to provide their opinions or ratings about the road safety impact of driver traits and external factors in different realistic driving scenarios.

A survey about factors affecting HGV Driving

Information for Participants (Ethics Ref: CS-2020-R9)

Purpose of this study:
We have developed a risk assessment system for HGV driving that automatically evaluates the risk associated with driving behaviours and environmental conditions. The purpose of this study is to evaluate the outputs of our system using stakeholders in the HGV industry i.e. drivers, transport managers, road safety specialists and researchers.

Nature of participation:
This study is open to any individual over the age of 18, who has experience in HGV driving and road safety. Participation is voluntary and relies on the participant completing a 5-10 minutes questionnaire. No identifiable information is collected in the study.

Benefits and risks of the research:
Your participation will help us evaluate the effectiveness, fairness and usefulness of our system, and provide the necessary feedback to improve our system. Our system aims to assist in developing effective real-time monitoring and feedback systems to assist drivers, providing important information about the impact of external factors on road safety to improve traffic laws, and ultimately, reduce the number of road incidents. The study has minimal risk for participation and data analysis.

Contact details of the ethics committee. If you wish to file a complaint or exercise your rights you can contact the Ethics Committee at the following address: cs-ethicsadmin@cs.nott.ac.uk

By clicking 'Next', I confirm that I have read and understood the study information, and I consent and wish to proceed.

* Required

Overview

In this study, we present user stories to participants that illustrate different driving scenarios. For each user story, we will ask participants to provide their opinions about the level of risk. The driving scenarios can be rated from 'very low' risk to 'very high' risk as shown in the figures below.

Please note that some user stories display aggressive or reckless driving styles. This does not imply in any way that HGV drivers are reckless. We simply wish to evaluate the effectiveness of our system in assessing extreme driving scenarios.

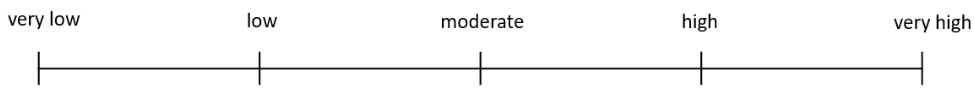
1. What is your occupation? *

Mark only one oval.

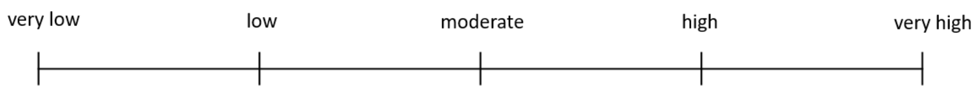
- Transport Manager
- Road Safety/Traffic Officer
- HGV driver
- Researcher/Academics
- Other: ______________________

Baseline User Story: A HGV driver is travelling on a sunny motorway with traffic moving at a fast pace. The driver maintains a speed below the maximum speed limit. The driver is attentive and well rested.

All participants considered the baseline driving scenario as 'Very Low to Low' risk.



User Story 1: A HGV driver is approaching a busy roundabout in a foggy weather. The driver reduces their speed well below the maximum speed limit of the road. The driver is attentive and well rested.



2. Q1.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

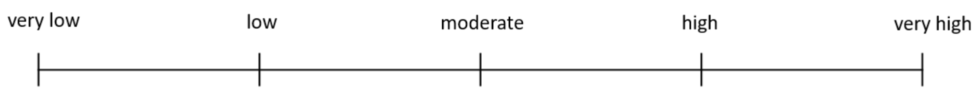
- [ ] Very Low
- [ ] Very Low to Low
- [ ] Low
- [ ] Low to Moderate
- [ ] Moderate
- [ ] Moderate to High
- [ ] High
- [ ] High to Very High
- [ ] Very High

3. Q1.b. Which factors were influential in your assessment? *

Check all that apply.

- [ ] Driving style
- [ ] Affective state
- [ ] Level of distraction
- [ ] Driving style, Affective state and Level of distraction
- [ ] Weather condition
- [ ] Traffic congestion
- [ ] Road Type
- [ ] Road type, Weather condition and Traffic congestion
- [ ] All of the above

User Story 2: A HGV driver is travelling in a rainy undivided highway with traffic moving at a fast pace. The driver has a total of 20 seconds driving at the maximum speed limit within 1 mile of the road. The driver is well rested but distracted.



4. Q2.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

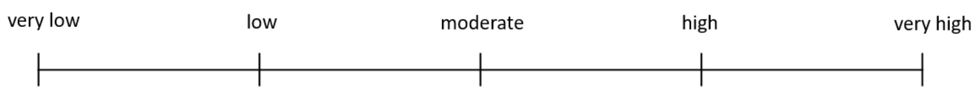
- Very Low
- Very Low to Low
- Low
- Low to Moderate
- Moderate
- Moderate to High
- High
- High to Very High
- Very High

5. Q2.b. Which factors were influential in your assessment? *

Check all that apply.

- Driving style
- Affective state
- Level of distraction
- Driving style, Affective state and Level of distraction
- Weather condition
- Traffic congestion
- Road Type
- Road type, Weather condition and Traffic congestion
- All of the above

User Story 3: A HGV driver is approaching a clear roundabout in a sunny weather. The driver maintains their speed. The driver is attentive but has a negative mood.



6. Q3.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

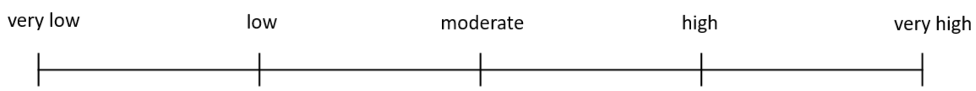
- [ ] Very Low
- [ ] Very Low to Low
- [ ] Low
- [ ] Low to Moderate
- [ ] Moderate
- [ ] Moderate to High
- [ ] High
- [ ] High to Very High
- [ ] Very High

7. Q3.b. Which factors were influential in your assessment? *

Check all that apply.

- [ ] Driving style
- [ ] Affective state
- [ ] Level of distraction
- [ ] Driving style, Affective state and Level of distraction
- [ ] Weather condition
- [ ] Traffic congestion
- [ ] Road Type
- [ ] Road type, Weather condition and Traffic congestion
- [ ] All of the above

User Story 4: A HGV driver is travelling on a sunny undivided highway with traffic moving at a moderate pace. The driver has a total of 6 harsh braking incidents within 1 mile of the road. The driver is distracted and frustrated by their in-vehicle delivery management system.



8. Q4.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

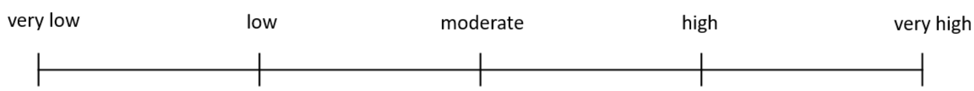
- [ ] Very Low
- [ ] Very Low to Low
- [ ] Low
- [ ] Low to Moderate
- [ ] Moderate
- [ ] Moderate to High
- [ ] High
- [ ] High to Very High
- [ ] Very High

9. Q4.b. Which factors were influential in your assessment? *

Check all that apply.

- [ ] Driving style
- [ ] Affective state
- [ ] Level of distraction
- [ ] Driving style, Affective state and Level of distraction
- [ ] Weather condition
- [ ] Traffic congestion
- [ ] Road Type
- [ ] Road type, Weather condition and Traffic congestion
- [ ] All of the above

User Story 5: A HGV driver is travelling in a foggy motorway with traffic moving at a slow pace. The driver produces a total of 1 harsh braking incidents within 1 mile of the road. The driver is distracted and in a negative mood.



10. Q5.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

- [ ] Very Low
- [ ] Very Low to Low
- [ ] Low
- [ ] Low to Moderate
- [ ] Moderate
- [ ] Moderate to High
- [ ] High
- [ ] High to Very High
- [ ] Very High

11. Q5.b. Which factors were influential in your assessment? *

Check all that apply.

- [ ] Driving style
- [ ] Affective state
- [ ] Level of distraction
- [ ] Driving style, Affective state and Level of distraction
- [ ] Weather condition
- [ ] Traffic congestion
- [ ] Road Type
- [ ] Road type, Weather condition and Traffic congestion
- [ ] All of the above

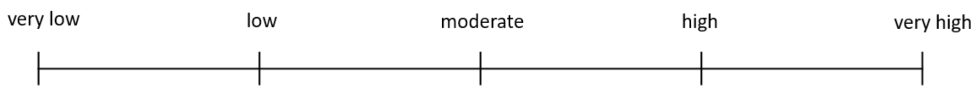
1. What is your occupation? *

Mark only one oval.

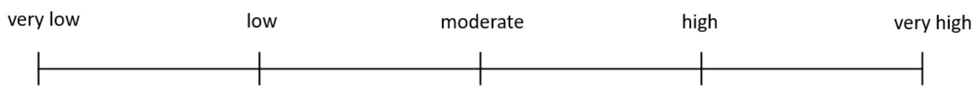
- Transport Manager
- Road Safety/Traffic Officer
- HGV driver
- Researcher/Academics
- Other: ______________________

Baseline User Story: A HGV driver is travelling on a sunny motorway with traffic moving at a fast pace. The driver maintains a speed below the maximum speed limit. The driver is attentive and well rested.

All participants considered the baseline driving scenario as 'Very Low to Low' risk.



User Story 1: A HGV driver is travelling on a icy rural road with traffic moving at a moderate pace. The driver produces a total of 2 rash overtaking incidents within 1 mile of the road. The driver is attentive but angry due to pressure from their manager to complete a job.



2. Q1.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

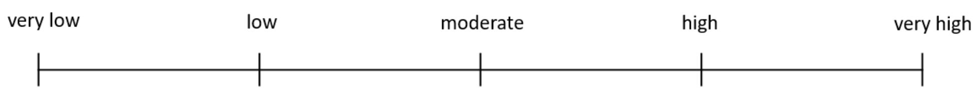
- [ ] Very Low
- [ ] Very Low to Low
- [ ] Low
- [ ] Low to Moderate
- [ ] Moderate
- [ ] Moderate to High
- [ ] High
- [ ] High to Very High
- [ ] Very High

3. Q1.b. Which factors were influential in your assessment? *

Check all that apply.

- [ ] Driving style
- [ ] Affective state
- [ ] Level of distraction
- [ ] Driving style, Affective state and Level of distraction
- [ ] Weather condition
- [ ] Traffic congestion
- [ ] Road Type
- [ ] Road type, Weather condition and Traffic congestion
- [ ] All of the above

User Story 2: A HGV driver is travelling in a snow urban road with traffic moving at a moderate pace. The driver reduces their speed well below the maximum speed limit and maintains an appropriate distance from preceding vehicles. The driver is attentive and well rested.



4. Q2.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

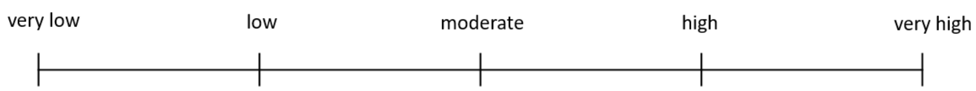
- [ ] Very Low
- [ ] Very Low to Low
- [ ] Low
- [ ] Low to Moderate
- [ ] Moderate
- [ ] Moderate to High
- [ ] High
- [ ] High to Very High
- [ ] Very High

5. Q2.b. Which factors were influential in your assessment? *

Check all that apply.

- [ ] Driving style
- [ ] Affective state
- [ ] Level of distraction
- [ ] Driving style, Affective state and Level of distraction
- [ ] Weather condition
- [ ] Traffic congestion
- [ ] Road Type
- [ ] Road type, Weather condition and Traffic congestion
- [ ] All of the above

User Story 3: A HGV driver is travelling on a sunny rural road with traffic moving at a fast pace. The driver produces a total of 1 harsh acceleration incidents within 1 mile of the road. The driver is in a positive mood but distracted.



6. Q3.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

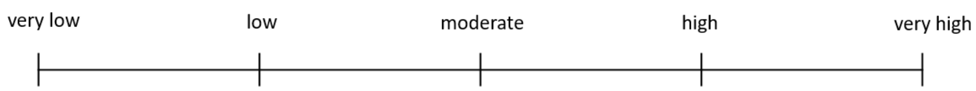
- [ ] Very Low
- [ ] Very Low to Low
- [ ] Low
- [ ] Low to Moderate
- [ ] Moderate
- [ ] Moderate to High
- [ ] High
- [ ] High to Very High
- [ ] Very High

7. Q3.b. Which factors were influential in your assessment? *

Check all that apply.

- [ ] Driving style
- [ ] Affective state
- [ ] Level of distraction
- [ ] Driving style, Affective state and Level of distraction
- [ ] Weather condition
- [ ] Traffic congestion
- [ ] Road Type
- [ ] Road type, Weather condition and Traffic congestion
- [ ] All of the above

User Story 4: A HGV driver is travelling on a snowy urban road with traffic moving at a fast pace. The driver produces no incidents within 1 mile of the road. The driver is attentive but frustrated by the amount of jobs assigned to them.



8. Q4.a. What level of risk will you consider the above driving scenario? *

Check all that apply.

- [ ] Very Low
- [ ] Very Low to Low
- [ ] Low
- [ ] Low to Moderate
- [ ] Moderate
- [ ] Moderate to High
- [ ] High
- [ ] High to Very High
- [ ] Very High

9. Q4.b. Which factors were influential in your assessment? *

Check all that apply.

- [ ] Driving style
- [ ] Affective state
- [ ] Level of distraction
- [ ] Driving style, Affective state and Level of distraction
- [ ] Weather condition
- [ ] Traffic congestion
- [ ] Road Type
- [ ] Road type, Weather condition and Traffic congestion
- [ ] All of the above