Developing Interpretable Predictive Models of Driver Situational Awareness in Conditionally Automated Driving

Sami Park

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

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Linda Ng Boyle, Chair

Ashis Banerjee, Chair

Prashanth Rajivan

Program Authorized to Offer Degree:

Industrial and Systems Engineering

University of Washington

Abstract

Developing Interpretable Predictive Models of Driver Situational Awareness in Conditionally Automated Driving

Sami Park

Co-Chairs of the Supervisory Committee:
Co-chair Linda Ng Boyle

Co-chair Ashis Banerjee
Industrial and Systems Engineering

Situational awareness (SA) among human drivers is important to understand for the advancement of vehicle automation. In conditionally automated driving scenarios, where the vehicle may approach its operational limits, the driver will be required to resume control within an appropriate time frame. This would suggest that real-time information about surroundings needs to be provided to the driver in a meaningful way while ensuring they are continually aware of their road environment.

The goal of this dissertation is to develop real-time predictive models of driver SA, focusing not only on performance but also on interpretability, providing users with insight into the driving context. The pursuit of this goal involves a structured approach, encompassing the following key steps: 1) Identification of driver-specific and environmental predictors, 2) Exploration of the relationships between three distinct levels of SA and the identified predictors, 3) Development of SA predictive models with various feature selection methods, and 4) Comparative assessment of the performance and interpretability of the developed models.

To accomplish the initial two objectives and collect data for the subsequent steps, two driving simulator studies were conducted with 40 and 56 participants respectively. These studies provide insights on the predictors needed for the real-time predictive models as well

as the complex relationships among different levels of SA (perception, comprehension, and projection) and the predictors. The data obtained from these studies serve as the foundational resource for objective 3: the development of real-time predictive models and Objective and 4: comparative analyses of the models in pursuit of the research goal.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number Page

LIST OF TABLES

Chapter 1
INTRODUCTION

1.1 Motivation

1.1.1 Driver situational awareness

Understanding the importance of situational awareness (SA) among human drivers becomes increasingly crucial as vehicle automation continues to advance. This heightened significance arises from the need for drivers to remain aware of their current and future driving conditions, especially in situations where they might need to take control of the vehicle. It is not always guaranteed that drivers will fully grasp the situation and react promptly. Over 94% of traffic accidents in the United States result from driver errors, and 40% of these errors are linked to recognition failures [118]. These recognition errors often stem from deficiencies in SA, which is precisely defined as "the perception of environmental elements and events in relation to time and space, the comprehension of their meaning, and the projection of their future status" [29]. SA serves as a key bridge between sensing information in the driving environment and making decisions that guide action [85].

Advanced driver assistance systems (ADAS) have emerged as a potential solution to enhance drivers' awareness of their surroundings, ultimately contributing to reduced accident rates [74, 120]. However, a balance must be maintained to prevent overwhelming drivers with excessive alerts. An overabundance of information can increase drivers' cognitive load, jeopardizing their ability to effectively allocate attention to safety-critical situations [12, 101]. This equilibrium is particularly critical for SAE level 2 systems, which heavily rely on human operators' capacity to perceive, comprehend, and anticipate driving scenarios. Recognizing that SAE level 2 systems will remain in use for some time, it is imperative to identify the factors that might negatively affect drivers' SA.

Additionally, even in SAE level 3 systems, the vehicle might reach its operational limit during adverse driving conditions, in which it would request the driver to take over the driving task within a specific time budget [6]. This emphasizes the importance of better understanding and promptly assessing driver SA [131, 23]. It is important for the driver to maintain a good level of SA and quickly regain it when necessary to safely negotiate the driving scenarios.

Takeover performance measures, including takeover time and takeover quality, are also associated with SA. For example, a low level of SA was associated with a longer takeover time, worse takeover decisions, and worse non-driving-related task (NDRT) performance compared to a high level of SA [144]. Systems for predicting and monitoring driver SA enable appropriate interventions to be introduced to improve takeover performance in conditionally automated driving [26].

1.1.2 SA evaluation methods

Researchers have adopted various approaches to identify and evaluate driver SA, incorporating both direct and indirect methods. The direct approach involves techniques such as the Situation Awareness Global Assessment Technique (SAGAT), which briefly interrupts ongoing situations to prompt drivers to share insights about their awareness [38]. On the other hand, indirect measures make use of diverse task performance indicators, including response times, accuracy, and physiological data such as eye-tracking, cardiovascular metrics, and electroencephalography [140, 142, 90, 80].

Query-based methods are often used to assess drivers' awareness of the current driving situation. The methods can be categorized as online or offline depending on whether the tasks are interrupted or not during the queries. In the offline approach, the ongoing tasks are paused or frozen and in the online approach, the queries are asked concurrently as an additional task without pausing the ongoing task. The offline approach allows for more in-depth probes compared to the online approach, given the pauses for queries.

SAGAT is the widely used offline method [38], where simulations of representative tasks or scenarios are paused at randomly selected times. SAGAT measures objective SA at these pauses by querying only the observed situation, unlike measures such as the Situational Awareness Rating Technique (SART), which solicits subjective ratings [40]. Pausing randomly makes it difficult for subjects to prepare for the queries and as a result, only partial information related to the query may be retained in working memory [30].

1.1.3 Related work

There have been many attempts to develop predictive models of driver SA from past studies that can be integrated into autonomous vehicles. These models employed features related to environmental characteristics such as object location and object type, and driver's eye movements to estimate driver SA.

For example, [109] used environmental complexity data from a driving simulator study in their SA predictive model. [144] used scene density and color differences between an object and the background to predict SA. Eye-tracking data was also used for the prediction during takeover transitions in conditionally automated driving. The study found that a greater number of fixations on the rear-view mirror and longer fixation durations were associated with better SA. [65, 66] conducted correlation and regression analyses between eye-tracking measures and driver SA and identified some promising variables that feature both spatial and temporal aspects of driver eye-glance behavior relative to objects of interest.

Yet, there are limitations of these models as well. For instance, many studies categorized environmental complexity into just two levels: low and high. Additionally, these previous studies have focused on predicting SA level 1 only and did not comprehensively include predicting SA levels 2 and 3. However, Levels 2 and 3 of SA are equally important as they account for understanding the road and projecting future conditions, respectively.

1.1.4 The need of comprehensive SA model

Past predictive models of driver SA can be enhanced by 1) incorporating more diverse factors and 2) considering different levels of SA in model development (especially levels 2 and 3), not only level 1. Driver SA encompasses various aspects, including the driver's perception of environmental elements and events, their understanding of the situation's meaning, and their capacity to predict future developments [29]. These three levels of SA do not necessarily follow a linear progression [34].

A model of SA should consider the varying complexities often observed in dynamic driving environments. Past SA models have considered visual differences (e.g., color, number of objects in the environment) [145] or environmental characteristics in terms of a binary outcome (easy, complex) [109] only. However, a more comprehensive SA model should also be able to account for the variation in visual complexity of objects that impact road safety (bicyclist, pedestrian, moving vehicles) as well the visual clutter in the background (traffic density, mixed mode traffic) that may impact driver's performance and attention [116, 25]. Increased visual clutter has been shown to increase search time and error rates for driving-related information [55, 17]. This can be compounded in urban environments where there is greater environmental complexity [96]. That said, the lack of information or environmental cues may also result in drivers being less aware of their surroundings [64].

Driving performance measures serve as important indicators of driver behavior and can provide valuable insights into driver SA. Commonly measured driving performance measures include driving speed, lane position, reaction time, and lateral and longitudinal vehicle control [14, 97]. For instance, reduced driving speed has been linked to distracted driving, and steering wheel metrics have been used to evaluate drivers' corrective actions [59].

Moreover, previous studies have explored the correlation between driver SA and other factors, such as environmental characteristics, driver eye-gaze behavior, characteristics of objects of interest (OOIs), and working memory [65, 66, 145, 136, 99]. Pedestrian behavior is dependent on the environment where they may not necessarily scan the environment [106]. Further, pedestrians may have cues that are not always recognized or appropriately detected by the human or system operator [41]. Building on this foundation, this work seeks to expand the predictors by incorporating environmental properties, particularly focusing on pedestrian interactions with the ego car and driving performance measures, to develop a more generalized and comprehensive SA predictive model.

1.1.5 Model Interpretability

Interpretability is inherently domain-specific, and as such, there is no universal definition [42, 58, 113]. Typically, an interpretable machine learning model is constrained in its form to be useful to someone or to adhere to structural knowledge within the domain. This structural knowledge may involve constraints such as monotonicity [49], causality, structural (generative) constraints, additivity [73], or physical constraints derived from domain expertise. Interpretable models might leverage case-based reasoning in complex domains [112].

Interpretability of models varies. Generally, rule-based, logistic regression, and decision tree models are easier to interpret compared to models like neural networks, Gradient Boosting Machine (boosted tree), or random forest [112].

In the context of driving, where decisions carry high stakes, interpretability is not only assessed in terms of accuracy, but also understanding the functionality of the model. Model interpretability in driving provides real-time SA information to drivers, offering explanations for the vehicle's behavior. For instance, if SA is assessed as low, the driver can receive an alert, accompanied by an explanation of why the vehicle perceives the situation as such. This approach enhances the overall driver experience and interaction with the vehicle.

The practical benefits of interpreting a model are as follows.

1. Providing drivers not only with the state but also the reasoning behind it (transparency effect) is crucial. Unreliable automation can impact the operator's perception and performance, and transparency has the potential to mitigate these effects [130].

2. Understanding the model allows for insights into factors causing lower SA, leading to improvements in the design of Advanced Driver Assistance Systems (ADAS). For instance, if a specific feature used for prediction is problematic, an explanation can help identify and address issues in the system.

1.1.6 Interpretability vs. Explainability

Researchers have advocated for 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' [112]. According to [48], explainable AI is defined as 'AI systems that can explain their rationale to a human user, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future.' For example, explainable AI in driving provides explanations for the vehicle's behavior after making decisions such as route choices or giving alerts.

While there is hope that creating methods to explain black box models will alleviate some of these problems, trying to explain black box models, instead of creating models that are interpretable from the start, is likely to perpetuate bad practices and could potentially cause catastrophic harm to society. To address this issue, researchers have advocated for designing models that are inherently interpretable [112].

In this thesis, interpretability will be used as a term to refer to both interpretability and explainability, as this work aims to first interpret models when possible and then explain the models if the inherent interpretation of the model is challenging.

1.2 Research Objectives

The current research gap underscores a deficiency in comparing the trade-offs between accuracy and interpretability in predictive models, particularly in the context of driving data. Little emphasis has been placed on interpretability, which is a critical factor in high-stakes decision-making. To address these gaps, the following objectives were set (see Figure 1.1).

The first objective focuses on the identification of comprehensive predictors of driver SA. It involves factors derived from eye-trackers, the road environment including the pedestrian-vehicle interactions, the vehicle, and the driver. Studies 1 and 2 collectively address this objective. Study 2 builds on the findings of Study 1, employing a broader range of features, including driving performance measures.

The second objective delves into investigating relationships between the three levels of

SA (perception, comprehension, and projection) and predictors. It seeks to understand the distinct dynamics of predictors and models across three levels of SA. Study 2 addresses this objective by collecting driver SA answers on each SA level respectively.

The third objective centers on the development of various SA predictive models. This objective utilizes data sets collected from Study 2. Various models were developed including logistic regression, decision tree models, Gradient Boosting Machine (boosted tree), random forest and shallow neural networks.

The fourth objective is comparative assessment of the performance and interpretability of the developed models. It involves the comparison between models designed with inherent interpretability and those explained post-hoc using approaches such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP).

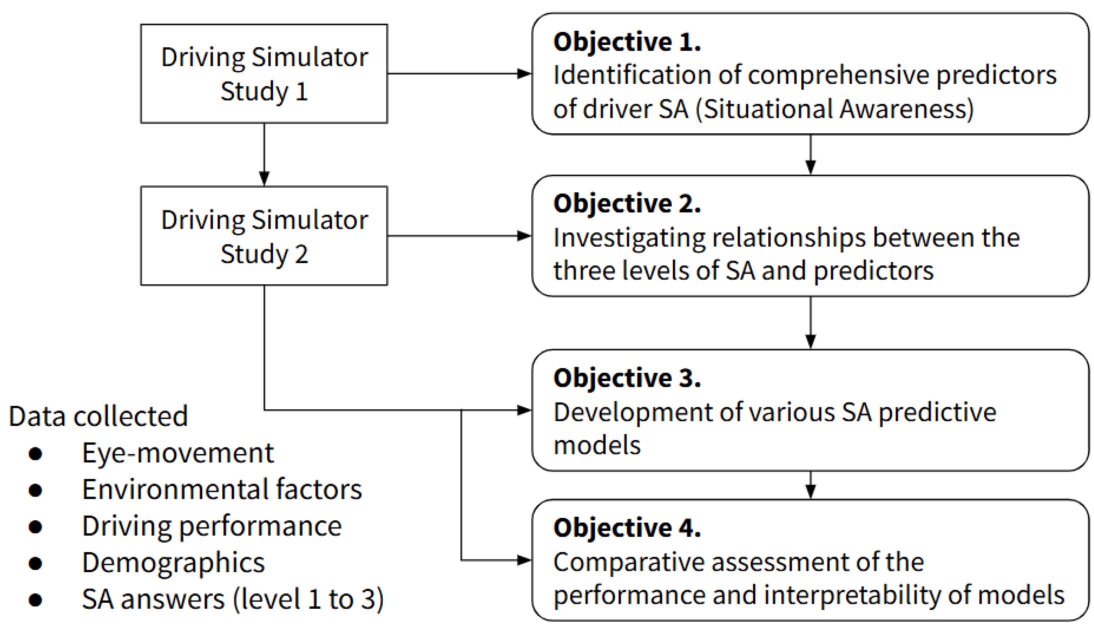


Figure 1.1: Illustration for research objectives

1.3 Organization of Thesis

Chapter 2 provides a comprehensive overview of Study 1, which aimed to identify predictors of driver SA, with a specific focus on environmental properties. The goal was to discern the factors influencing driver SA in relation to the surrounding environment.

In Chapter 3, the focus shifts to Study 2, which builds upon the findings of Study 1. Study 2 has two main objectives: 1) to identify additional factors contributing to driver SA, including more detailed environmental properties, driving performance measures, and pedestrian interaction factors; and 2) to investigate the relationships between the three levels of SA and these predictors. This study involves a nuanced exploration of how various factors impact different levels of driver SA, providing a deeper understanding of the dynamics involved.

Chapter 4 delves into a detailed discussion of Objective 3, outlining the research methodology designed to develop predictive models using various feature subsets. This chapter serves as a crucial bridge between the initial identification of predictors and the subsequent development of predictive models for driver SA.

Chapter 5 presents the final phase of the research, focusing on evaluating and comparing the performance and interpretability of the driver SA predictive models. Using standard metrics such as accuracy, precision, recall, and F1 score, model effectiveness was assessed. SHAP values were applied to evaluate interpretability, offering insights into both local predictions and global model behavior. Recognizing the inherent differences in model architectures, this chapter emphasizes a nuanced evaluation of interpretability across diverse models to ensure a comprehensive understanding of their decision-making processes.

Finally, Chapter 6 serves as a summary of the research, providing a consolidated view of the research journey. It encapsulates the key findings from the initial studies and outlines subsequent research directions, methodological intricacies, and results.

Chapter 2

STUDY 1: IDENTIFICATION OF PREDICTORS

2.1 Introduction

The objective of this study is to identify comprehensive predictors of driver SA. This is achieved using data collected from a human subject driving simulator study [136, 99]. The study computed a scenario-wise SA measure that is used in a model to identify the impacts of environmental features and visual complexity on SA scores, which is then compared to the object-wise perspective presented by [136]. This study was approved by the University of Washington Institutional Review Board (IRB) Committee (No: STUDY00013730).

2.2 Method

2.2.1 Participants

Participants were recruited using email lists at the University of Washington and from online sources (e.g., Facebook and Craigslist) around the Seattle, WA area. The participants were required to have a valid US driver's license, drive at least 3,000 miles a year, and have normal or corrected-to-normal visual acuity.

Drivers were screened into two groups based on age, 18 - 40 years old (younger), and over 40 years old (older). Sign-up for the experiment was conducted on a first-come-first-serve basis. Once all the slots for each age group were filled, no additional participants were selected. In total, 40 drivers were selected to participate (Table 2.1).

There were 23 drivers with normal eyesight and 17 that required glasses or contact lenses on a daily basis (diopter within the range of -5.0D to +3.0D). Upon successful completion, the drivers were compensated US$37.50 for their time in the study.

Table 2.1: Summary Statistics of Study Participants

			Younger (18-40)		Older (40+)	
Variable		Total	Female	Male	Female	Male
Subjects (n)		40	10	10	10	10
Age	range (yrs)	18 - 61	18 - 40	20 - 33	41 - 60	41 - 61
	mean (yrs)	36.7	22.7	24.0	49.3	50.9
	SD (yrs)	14.9	6.5	3.6	6.6	8.0
Driving experience	range (yrs)	2 - 45	2 - 24	4 - 11	16 - 44	24 - 45
	mean (yrs)	19.1	6.4	6.8	31.3	33.2
	SD (yrs)	14.5	6.6	2.3	7.8	8.4

2.2.2 Simulator Environment

Driving simulators are widely used for studies on driver behavior and performance as they encounter at-risk situations that cannot be safely tested on the road [3, 133, 94]. This study used the hardware of a fixed-based miniSim[1] driving simulator, featuring real-world driving scenes displayed across three 42-inch plasma displays with a resolution of 1,280 × 720. The setup was designed to be comparable with other immersed driving simulator environments in terms of manual driving demand, participants' gaze behavior, and awareness of road hazards [65]. To approximate manual driving demands, drivers were asked to operate the steering wheel and pedals in the driving simulator to match the movement of the vehicle in the video.

2.2.3 Driving Scenarios

The videos used were recorded in urban environments under daylight conditions in San Francisco, California, with a human driver using an instrumented vehicle equipped with various sensors [104] on both rural and urban roads. The videos presented various densities of moving

[1]https://www.nads-sc.uiowa.edu/minisim/

objects of interest that include vehicles, pedestrians, and cyclists. The visual complexity of the 2D scenes was calculated as the ratio between the compressed and uncompressed image file sizes using the following equation:

$$Visual\ complexity = \frac{Compressed\ size}{Original\ size} \tag{2.1}$$

where the compressed image size was calculated based on the zip image compression algorithms with deflate compression using the R function `img_complexity()` from package `imagefluency` (Version 0.2.3) [86]. Theoretically, this value ranges from zero (no visual complexity) to one (highest visual complexity). In this study, the visual complexity of scenario images ranged from 0.2 to 0.7 (Mean = 0.43, SD = 0.10).

The compression rate was used as an approximation of the collection of various environment features. For example, the road user's ability to accurately perceive pedestrian behavior is assumed to be impacted by the visual clutter of the buildings or the scene contrast given shadows. Scenarios with greater visual complexity tend to have more road users and buildings with greater contrast between objects.

Event videos were classified into five levels by visual complexity: [0.2,0.3), [0.3-0.4), [0.4-0.5), [0.5-0.6) and [0.6,0.7). Example scenarios are shown in Table 2.2. The proportion of the visual complexity level of the selected videos was 9.85%, 28.5%, 35.8%, 21.2%, 3.98%. The images with a complexity level outside this range (less than 0.2 or greater than 0.7) accounted for less than 1% of all scenes and were removed from the eligible video pool.

2.2.4 Capturing SAGAT Responses

A tablet personal computer (13-inch iPad Pro) was used to record participants' SAGAT responses (see Figure 2.1). The tablet PC display would first show the same driving scene as the scenario displayed in the simulator with objects of interest (OOI) purposely removed by professional photo editors using Photoshop. The OOIs included 1) pedestrians, vehicles, and cyclists on the road, 2) vehicles that may potentially move, 3) pedestrians close to the road, and 4) pedestrians, cyclists, and vehicles that are on the same block as drivers. Non-OOI (e.g., vehicles parked on the side of the road) were considered less risky compared to OOIs and were not removed from the touch recorder image.

When the simulator scenario was paused, drivers were instructed to pick up the tablet PC and use a stylus to mark the locations and types of every detected Object of Interest (OOI) observed in the scene. A yellow circle would appear anytime a participant touched any location on the scene displayed on the iPad. This provided confirmation to participants that an object was selected. The participant would then choose one of three buttons to select the type of object (pedestrian, vehicle, or cyclist) they will be marking. Once the button is selected, the yellow circle would change to a green circle, red box, or blue box for a pedestrian, vehicle, or cyclist, respectively. An undo button was used to correct a previous touch. When the drivers were done, they would press 'click when complete' to move to the next scenario. Between each scenario, a buffer screen was used to prevent drivers from predicting the next SAGAT. The touch recorder allowed spatially continuous data of driver SA to be captured as an x-y coordinate.

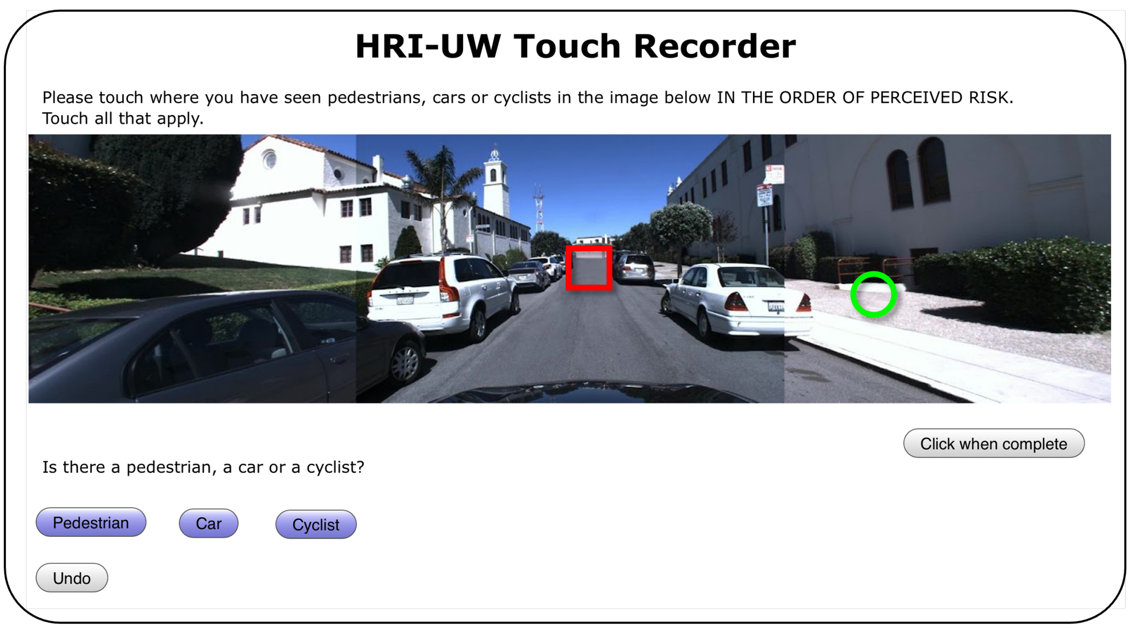
2.2.5 Experimental Design

A split-plot design was applied for this study. That is, the 40 participants were split into four whole plots (younger female, younger male, older female, and older male). Each of the whole plots included 10 subplots, which were defined as participants (see Table 2.3). Each whole plot was run as a balanced incomplete block design. There were 75 scenarios used in this study, which were split into five chunks with 15 scenarios per chunk. Each participant experienced two out of five chunks or 30 (2x15) scenarios. As a result, each scenario was replicated 16 times[2] for this study.

[2](40 participants $\times$ 30 scenarios)/75 scenarios

(a) Scene

(b) Empty scene - displayed on simulator and the touch recorder at pause



(c) Touch recorder - marked scene

Figure 2.1: Example displays from actual driving scene (a), after the scene paused (b), and as viewed on the touch screen recorder (c)

2.2.6 Procedure

Upon arrival, participants reviewed and signed the informed consent. The participants were then provided information on study expectations and instructions on the tasks they would conduct and how to use the touch recorder. The study consisted of three parts: 1) a pre-experiment survey, 2) a practice drive, and 3) the main study which included two study drives.

The pre-experiment survey included demographic information and driving experience (see Appendix A). The study utilized a wearable eye tracker, specifically the Tobii Pro Glasses 3[3]. Before the practice drive, the researcher would calibrate the eye-tracking glasses for each participant. For drivers who wore glasses, corrective lenses were applied to the eye tracker. Participants were informed before study enrollment that they will need to wear the eye tracker during the experiment. The eye tracker was lightweight (76.5 grams with cable) and similar to wearing sunglasses. That said, participants were informed that they can pause or stop if they felt uncomfortable.

Before the main study, a 5-minute practice drive was conducted to familiarize drivers with the driving simulator. In the main study, participants were instructed to operate the steering wheel, pedals, and turn signals in the driving simulator while watching real-world videos. They were asked to match the movement of the vehicle in the video. The participants' control of the vehicle provided no feedback to the driving scene but rather, was intended to immerse the participants in the driving scenarios.

In the main study, the video paused randomly during each scenario, preventing drivers from predicting when it would occur. At each pause, drivers picked up the iPad and answered SAGAT questions using the touch recorder. A 10-minute break was given between the study drives. Throughout the experiment, drivers were encouraged to ask questions about using the touch recorder at any time. Drivers were also informed that they can stop the experiment at any time if they feel sick or uncomfortable. Driving scenarios at pauses had various

[3]https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3

densities of road users and situations. Each study drive took about 20 minutes and the entire experiment was completed in about 75 to 90 minutes.

2.2.7 Measures

Dependent Variable

An SA score, represented as a percentage, was computed using the proportion of Objects of Interest (OOI) correctly identified by drivers in each scenario:

$$SA\ score\ (\%) = \frac{Number\ of\ OOI\ correctly\ identified}{Total\ number\ of\ OOI} \times 100\% \tag{2.2}$$

An answer was marked correct if the participant's touch position was located within the bounding box (see Figure 2.2). A touch position located within the bounding box but marked as the wrong type of object or touch outside the bounding box was considered a wrong answer. Wrong answers or no answers were not counted as a penalty in the SA score.

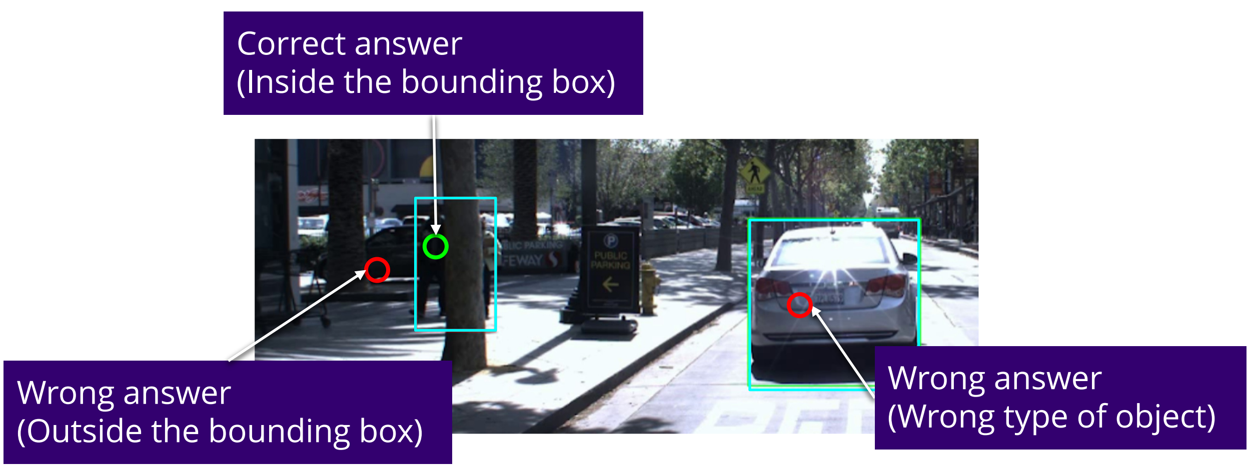


Figure 2.2: Illustration for labeling correct and incorrect answers

Independent Variables

The independent variables included fixed effects and a random effect for participants, accounting for the 30 scenarios completed by each individual. Duration, a fixed effect, is defined as the accumulated driving time (in minutes) from the start of the drive to the beginning of each SAGAT pause (Range (R): 0.7-17.8, Mean (M) = 8.81, Standard Deviation (SD) = 5.03). The remaining fixed effects were extracted for each scene at every pause and are defined as follows.

- Roadway type (2 levels): segment (68 %), intersection (32%).

  The intersection types included signalized, non-signalized, and roundabouts based on the definitions of the Federal Highway Administration (FHWA)[4]. Previous studies have shown that intersections [65], freeways [76], and various other road types [56] can affect drivers' ability to attend to safety-critical situations.

- Lighting (2 levels): bright (61%), less bright (39%).

  The 'bright' lighting scenario shows higher contrast and brightness compared to 'less bright' lighting scenario (see Figure 2.3). Drivers may find it harder to identify objects under less bright road conditions. In this study, all scenarios were recorded in the daytime. Darker scenarios were considered during pilot testing but were excluded because it was too difficult to detect objects even manually.

(a) Bright　　(b) Less bright

Figure 2.3: Scenarios with different lighting conditions

[4] `https://highways.dot.gov/safety/intersection-safety/safe-system-intersections`

- N_{object}: number of total objects, including OOIs and non-OOIs (R: 1-46, M = 17.41, SD = 7.66).

- $P_{ped+cycle}$: percent of pedestrians and cyclists out of all possible OOIs and non-OOIs (R: 0.05-1.00, M = 0.30, SD = 0.22). This variable provides a more complete comparison of vulnerable road users to cars than a variable that includes only pedestrians or only cyclists.

- P_{OOI}: percent of objects of interest out of all the objects including both OOIs and non-OOIs; this includes pedestrians, cyclists, and vehicles (R: 0.05-1.00, M = 0.34, SD = 0.22).

- $P_{OOI,peripheral}$: percent of objects of interest in peripheral view out of total OOIs (R: 0-0.50, M = 0.06, SD = 0.12). The peripheral view in this study refers to the view displayed on the left and right display (23.5 to 70.5 degrees of view on both sides) and outside the center display (47 degrees of view) as depicted in Figure 2.4.

- Visual complexity: the ratio of the compressed image size of the scenario and original size (R: 0.2-0.7, M = 0.43, SD = 0.10).

2.2.8 Data Processing and Analysis

There were 1,200 SA scores (40 participants × 30 scenarios) collected from this study. The distribution of SA scores is shown in Figure 2.5. Out of these, 754 SA scores equaled zero, indicating that no Objects of Interest (OOI) locations were correctly identified in the scenario.

A hurdle model was used to account for the excess zeros in the SA scores [20, 93]. The model consists of two parts; 1) a mixed-effects binary logit model that shows the likelihood that participants have a non-zero SA score (yes, no), and 2) a linear mixed-effects model to examine the likelihood of a higher positive, non-zero SA score [89]. Non-zero SA scores ranged from 6.67 to 100 (M=26.93, SD=17.12).

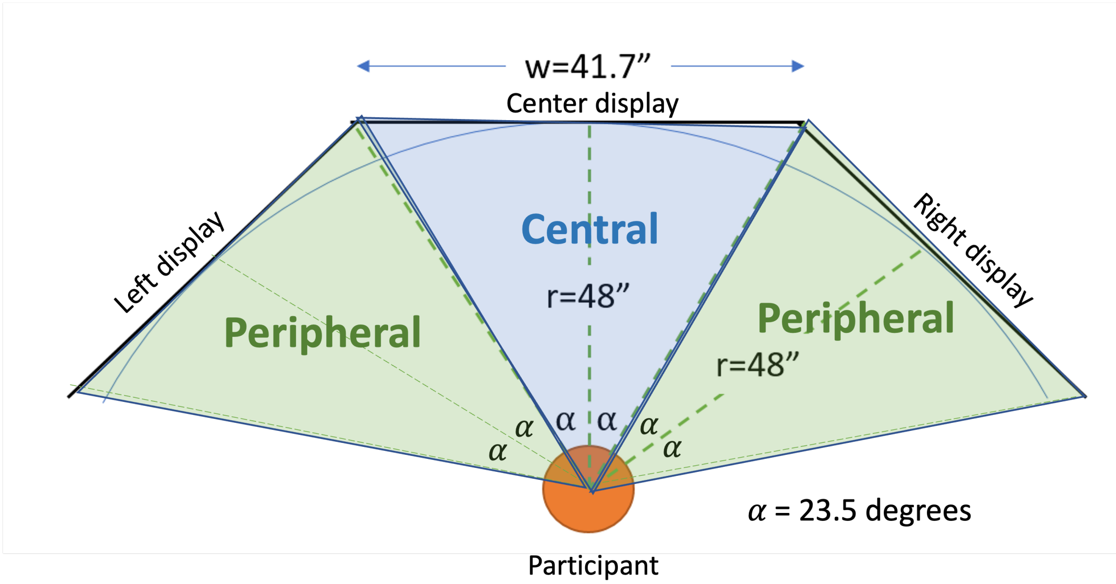


Figure 2.4: Top-down view of the simulator setting and the peripheral view

To address multicollinearity, the number of pedestrians, number of OOIs, and number of OOIs in peripheral view were transformed into ratios of total objects. The statistical software package `R v4.2.3` was used as the coding language for all data processing and analysis. Function `dredge` from package `MuMIn v1.47.5` in R was used for model selection. A full model that included all independent variables, their interaction terms, and the random effects was the initial starting point. From there, a backward stepwise approach was applied and each model was evaluated for goodness of fit based on the Akaike Information Criterion (AIC). Function `glmer` from package `lme4 v1.1-32` was used for building the model.

2.3 Results

2.3.1 Hurdle model – Mixed-effects binary logit portion

The final mixed-effects binary logit model is shown in Table 2.4. All variables and interaction terms included in the final model except $P_{OOI,peripheral}$ had a significant impact on the SA score – whether the SA score was zero or not. Scenarios with shorter duration, higher

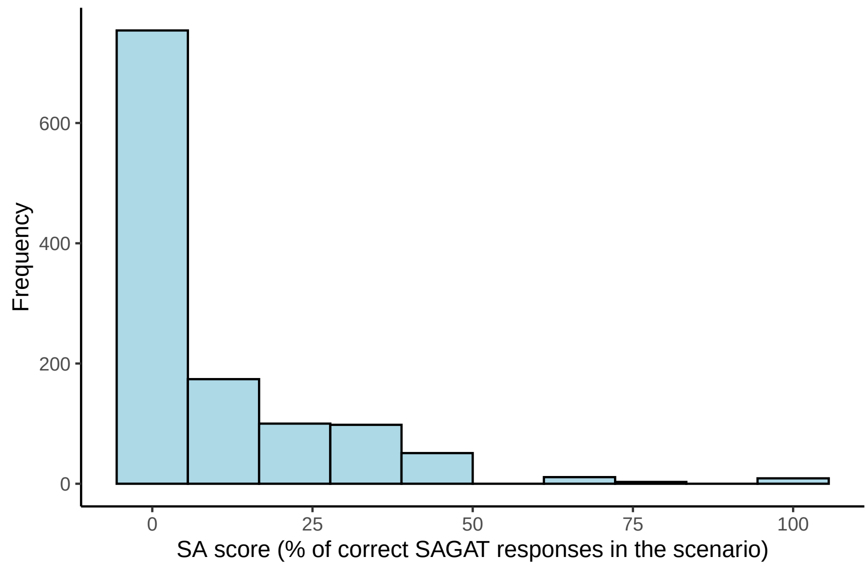


Figure 2.5: Distribution of SA scores

visual complexity, more objects, greater percentage of vulnerable road users (pedestrians and cyclists), and greater percentage of OOIs are more likely to have a non-zero SA score. A violin plot confirms that the likelihood of a non-zero is associated with higher visual complexity (see Figure 2.6). More specifically, the likelihood of a zero SA score is more spread out with a long tail toward the lower values. There are more non-zero values in the higher visual complexity levels. This is also confirmed in the heatmap shown in Figure 2.7.

In terms of interaction terms, the log-odds ratios corresponding to an increase in $P_{ped+cycle}$ of one percent for two P_{OOI} groups which differ by one percent will be $\exp(-3.17 \times 0.01) = 0.97$ of their current log-odds ratios. That is, if we increase the value of $P_{ped+cycle}$ by 1%, the log-odds ratios for two different groups of people (P_{OOI} groups) will change. The change will be such that the new log-odds ratios will be about 0.97 times the original log-odds ratios.

The log-odds ratios corresponding to an increase in visual complexity of 0.01 for two N_{object} groups which differ by one object will be $\exp(-0.31 \times 0.01) = 0.99$ of their current log-odds ratios. That is, if we increase visual complexity by 0.01 for two different groups of objects (N_{object} groups) that originally differ by one object, the log-odds ratios for these groups will change. The change will be such that the new log-odds ratios will be about 0.99 times the original log-odds ratios.

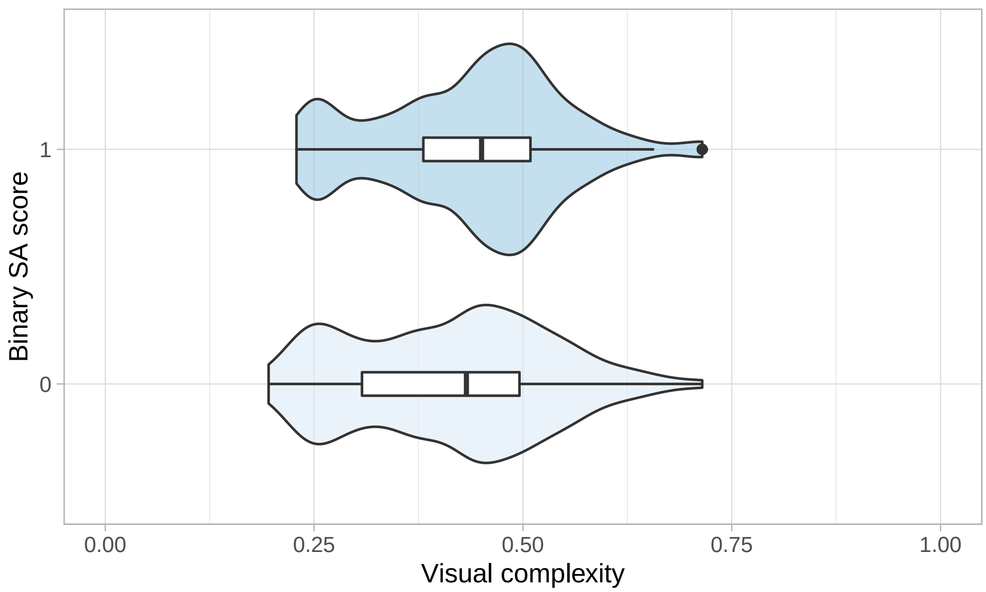


Figure 2.6: Violin plot of binary SA score given visual complexity

2.3.2 Hurdle model – Linear mixed-effects portion

The results of the linear mixed-effects model (Table 2.5) show that all main effects, except roadway type, had a significant impact on the SA score at the 0.05 level. Scenario-wise SA tends to be greater with lower visual complexity, lower number of objects ($N_{objects}$), and less bright lighting conditions. Greater SA was also observed when there was a smaller percentage of vulnerable road users ($P_{ped+cycle}$), a smaller percentage of OOIs (P_{OOI}), and a smaller percentage of OOIs in the peripheral view ($P_{OOI,peripheral}$). A 2D heatmap showed that lower visual complexity (less than 0.5) generated higher SA scores including some scores at 100% or close to 1 (see Figure 2.7).

In terms of the interaction terms, all were significant except two: 1) the interaction of lighting and roadway type and 2) the interaction of visual complexity and roadway type. The interaction terms provide insights on the relationships among the independent variables. For example, the model shows that each additional percent increase in $P_{ped+cycle}$ will increase the impact of P_{OOI} on the SA score by $0.56 \times 0.01 = 0.0056$. That is, the more vulnerable users there are in the scene, the stronger the impact of P_{OOI}. Each additional unit of N_{object} also increased the impact of $P_{OOI,peripheral}$ on the SA score by 0.03, and increased the impact of visual complexity on the SA score by 0.03. This suggests that the greater the number of objects in the scene, the stronger the impact of $P_{OOI,peripheral}$ and visual complexity.

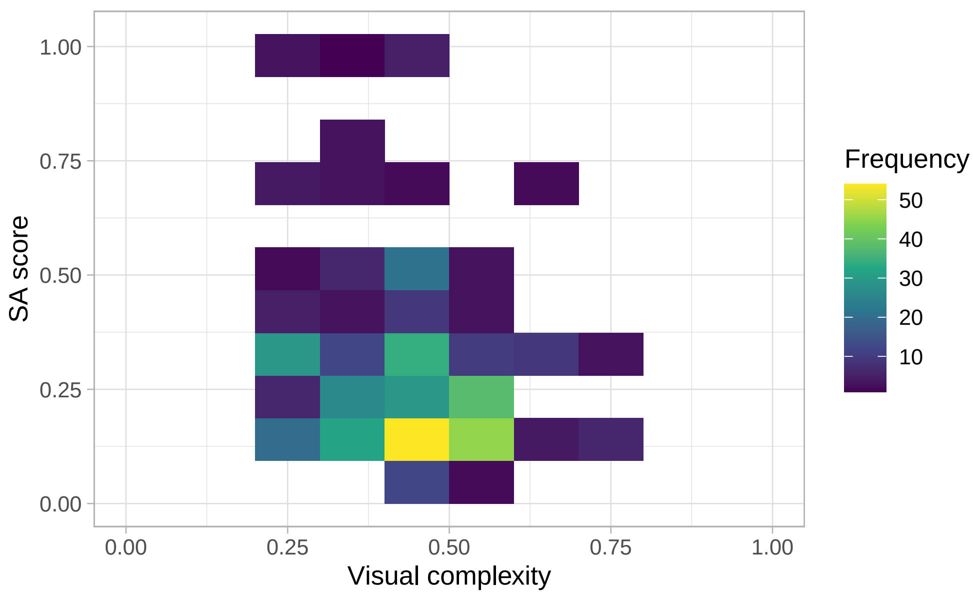


Figure 2.7: Heatmap of 2D bin counts for SA score given visual complexity

2.4 Discussion

Dynamic driving environments contain many environmental features that may impact the drivers' ability to attend appropriately to the roadway. Drivers often need to identify several objects of interest among different levels of visual clutter in the background. In this study, a hurdle model is used to predict driver SA. The objective measure of SA used in this study is based on the participant's selection from the entire driving scene. Hence, it does not force participants to place their visual experience into predetermined categories. The method used in this study builds upon other objective metrics of SA that use forced choices and discrimination tasks [65, 71]. The scenario-wise SA score was used as the outcome in a hurdle model to examine the impact of visual complexity, object density, lighting, and roadway type on drivers' SA.

The model findings suggest that higher visual complexity or more objects in the environment lead to a greater likelihood of a non-zero SA score but the SA score tends to be lower. That is, it is easier for drivers to perceive some objects in scenes with high visual complexity or more objects but will be more difficult for them to correctly identify a larger proportion of objects. Drivers tend to switch between "goal driven" and "data driven" information processing throughout their drive, which may impact their SA score [37]. Greater visual complexity or more objects in the scene corresponds to a greater amount of visual information and clutter in scenarios [28]. The findings showed that drivers are more likely to attend to cues that are most salient when there is too much information in the environment.

The visual complexity and total number of objects consist of two information types: useful OOIs and non-useful objects. The non-useful objects can be considered a distraction for drivers. Ideally, drivers can develop SA for all OOIs and filter out all the useless information. The percentage of OOIs sharing the same tendency with visual complexity and the number of total objects indicates that too much information, even if it is useful, can force drivers to focus only on the most salient objects. This poses challenges in developing driver assistance systems that can effectively filter out useless information while also assisting drivers in developing SA for missed OOIs.

A higher percentage of vulnerable road users (pedestrians and cyclists) was also associated with a higher likelihood of a non-zero SA score but the SA score was lower when compared to the percent of cars. It should be noted that a higher percentage of cars in our study does not necessarily indicate that the scene included only cars. In fact, the videos were selected from events with crossing pedestrians. This further emphasizes the need to help drivers improve

SA, especially when vulnerable road users are present.

Visual complexity was used to represent the visual clutter or the amount of visual information from the entire driving scene. Our model showed that it is a useful measure to identify drivers' SA. Because this measure was not strongly correlated to other measures (< 0.3), it can be used as an independent variable for SA prediction. Further, visual complexity and the number of objects provide different outcomes for each part of the hurdle model. For the binary model (whether SA score is 0 or not), the two variables cancel each other out. For the mixed linear model (SA scores), the two variable promotes each other's impact. This demonstrates that they are different measures that provide insights on both the useful and non-useful OOIs.

In the mixed-effects binary logit model portion of the hurdle model, the accumulated time from the start of a drive to a SAGAT pause was found to have a significant impact at the p-value 0.05 level. The negative coefficient suggests that as the duration increases, drivers were less likely to generate correct SA for the scenario. If there was a learning effect, one might expect driver SA to improve over time. However, the negative sign may indicate a potential limitation of the study, as it suggests that drivers may become less patient in answering the SAGAT questions as time passes.

Another study limitation is that SAGAT may not reflect actual SA for scenarios with a large number of objects and where drivers have to solely rely on their memory to mark objects [43, 50]. This may have been the case for the high-density traffic scenarios where drivers may not have been able to provide a response to all objects, not because they did not perceive them but because they may not have been able to retain everything in working memory. In addition, the videos used in this study were of scenes during daytime hours only. Drivers may use different visual cues during the nighttime.

The scenario-wise results were compared to the object-wise model presented in [136]. For both models, visual complexity was significantly associated with SA scores. The results indicate the robust impact of visual complexity on driver SA. This highlights the importance of visual complexity for predicting drivers' SA across various scales.

2.5 Conclusion

The goal of this study was to understand the impacts of environmental properties and visual complexity on drivers' situational awareness. The findings of this study identifies comprehensive environmental features that could be used to build predictive models of driver SA in complex driving situations. This information can be used to create systems that effectively assist drivers when needed.

Table 2.2: Examples of scenarios with different image visual complexity

Visual complexity	< 0.3	[0.3, 0.4)	[0.4, 0.5)	[0.5, 0.6)	> 0.6
Example scenarios					

Table 2.3: One whole plot contained 10 participants (subplots), and each subplot encountered two out of five chunk combinations

	Participants									
Drive	1	2	3	4	5	6	7	8	9	10
1st	3	4	2	2	1	1	4	5	3	5
2nd	5	1	4	3	3	2	5	1	4	2

Note: Each table cell refers to a chunk index (1 to 5).

Table 2.4: The likelihood of an SA score = 0 (yes, no) using mixed-effects binary logit model

	Estimate	S.E.	z value	Pr(> |z|)	Odds Ratio	95% CI
Intercept	-4.14 ***	0.67	-6.59	<0.001	0.01	(-5.73, -3.10)
Duration(minute)	-0.04 **	0.01	-3.08	0.002	0.96	(-0.07, -0.02)
Visual complexity	7.07 ***	1.52	4.96	<0.001	1803.04	(4.54, 10.46)
Object property						
N_{object}	0.15 ***	0.03	4.61	<0.001	1.16	(0.09, 0.22)
$P_{ped+cycle}$	1.87 **	0.69	2.70	0.007	6.50	(0.51, 3.23)
P_{OOI}	1.81 **	0.66	2.73	0.006	6.14	(0.51, 3.11)
$P_{OOI,peripheral}$	-1.08 ·	0.62	-1.73	0.084	0.340	(-2.30, 0.14)
Interaction terms						
$P_{ped+cycle}$:P_{OOI}	-3.17 *	1.37	-2.32	0.020	0.04	(-5.84, -0.49)
Visual complexity:N_{object}	-0.31 **	0.07	-4.31	<0.001	0.73	(-0.45, -0.17)

Number of Obs.: 1200 AIC: 1490.7 SD of Random effect: 0.54

Log Likelihood (convergence): -735.4 Log Likelihood (initial): -791.8

*** $p < 0.001$, ** $p < 0.01$, *$p < 0.05$, · $p < 0.1$

Table 2.5: Likelihood of higher SA scores using a linear mixed-effects model

	Estimate	S.E.	t value	Pr(> |t|)	95% CI
Intercept	1.24 ***	0.10	11.86	<0.001	(1.04, 1.45)
Visual complexity	-0.89 ***	0.23	-3.80	<0.001	(-1.35, -0.43)
Lighting					
Bright					
Less bright	-0.16 **	0.06	-2.81	0.005	(-0.28, -0.05)
Roadway type					
Intersection					
Segment	-0.11 ·	0.06	-1.74	0.083	(-0.23, 0.01)
Object property					
N_{object}	-0.03 ***	0.01	-5.58	<0.001	(-0.04, -0.02)
$P_{ped+cycle}$	-0.26 ***	0.07	-3.77	<0.001	(-0.39, -0.12)
P_{OOI}	-0.77 ***	0.07	-11.28	<0.001	(-0.91, -0.64)
$P_{OOI,peripheral}$	-0.61 ***	0.14	-4.29	<0.001	(-0.89, -0.33)
Interaction terms					
Visual complexity:N_{object}	0.03 **	0.01	2.72	0.007	(0.01, 0.05)
$P_{ped+cycle}$:P_{OOI}	0.56 ***	0.14	4.03	<0.001	(0.28, 0.82)
N_{object}:$P_{OOI,peripheral}$	0.03 ***	0.01	5.62	<0.001	(0.02, 0.04)
Lighting(less bright):Roadway(segment)	0.06 ·	0.03	1.87	0.062	(-0.00, 0.12)
Visual complexity:Roadway(segment)	0.14	0.13	1.08	0.280	(-0.12, 0.40)
Visual complexity:Lighting(less bright)	0.32 *	0.14	2.29	0.022	(0.05, 0.59)

Number of Obs.: 446 AIC: -505.6 SD of Random effect: 0.02

Conditional R^2: 0.50 Marginal R^2: 0.49

Log Likelihood (convergence): 268.9 Log Likelihood (initial): 154.9

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, · $p < 0.1$

Chapter 3

STUDY 2: EXPLORATION OF THE RELATIONSHIPS BETWEEN THREE DISTINCT LEVELS OF SA AND THE IDENTIFIED PREDICTORS

3.1 Introduction

The objective of this study is to investigate the factors that are most critical in predicting driver SA. This includes understanding drivers' ability to perceive, comprehend, and project pedestrian behavior across varying SA Levels (ranging from level 1 to 3). By exploring the relationship between driving performance measures, pedestrian interaction properties, and driver SA, this study aims to contribute to the development of efficient ADAS systems that prioritize critical alerts, enhance driver safety, and promote the successful integration of autonomous vehicles in real-world driving scenarios.

3.2 Method

3.2.1 Participants

A total of 56 participants took part in the study, with an equal distribution between those who self-identified as females and non-females. The participants ranged in age from 18 to 52, with an average age of 26. Participants were recruited through email lists at the University of Washington (WA) and online sources such as Facebook and Reddit in the Seattle, WA area. The inclusion criteria for participants were as follows: possessing a valid US driver's license, driving at least 3,000 miles per year, having normal or corrected-to-normal visual acuity, being fluent in English, and not having participated in a driving simulator study in the past 6 months to minimize learning biases. Upon successful completion, the drivers were compensated US$50 for their time in the study. This study was approved by the University of Washington Institutional Review Board (IRB).

3.2.2 Apparatus

Driving simulator

In this study, the hardware of a fixed-based miniSim driving simulator [125] was used with driving scenarios built using a custom Unreal Engine-based driving simulator for urban environments, replicating downtown Mountain View, CA. The hardware setup consisted of three 42-inch plasma displays, each with a resolution of 1,280 $\times$ 720. The simulator presented driving scenarios across the three screens, where each screen represented a distinct channel (left, central, and right) of the simulator scenario. The drivers were instructed to interact with the simulator by operating the steering wheel, accelerator pedal, and brake pedal, simulating the experience of driving a real car without using any advanced driver assistance functions.

Eye tracker

A wearable eye tracker (Tobii Pro Glasses 3) was used in the study. Before the practice drive, the researcher calibrated the eye-tracking glasses for each participant. Corrective lenses were applied to the eye tracker for drivers who wore glasses.

Input interface

A touch screen tablet (13-inch iPad Pro) was used by participants to record their SAGAT responses.

3.2.3 Experimental Procedures

Upon arrival, participants reviewed and signed an informed consent form. They received information about the study expectations and instructions for the tasks and the use of the touch recorder. The study consisted of three parts: (1) a pre-experiment questionnaire, (2) a practice drive, and (3) the main study, which included three study drives and a post-experiment questionnaire.

The pre-experiment questionnaire included demographic information (age, gender), driving experience (years of experience, miles driven/month), and the Driver Behavior Questionnaire (DBQ) [107, 84]. Prior to study enrollment, participants were informed that they would be required to wear an eye tracker during the experiment. Before the practice drive, the researcher calibrated the eye-tracking glasses for each participant. During the study, they were informed that they had the option to pause or stop if they felt uncomfortable at any point. To familiarize participants with the driving simulator, a 5-minute practice drive was conducted before the main study.

In the main study, participants were given instructions to operate the steering wheel and pedals in the driving simulator. Their task was to drive through intersections that were representative of a US urban downtown area, following the road without making any left or right turns. The simulator would pause randomly during each scenario so that prediction of the pause location would be minimized. At each pause, drivers would pick up the touch screen tablet and answer SAGAT questions. A 10-minute break was provided between the study drives. Participants were instructed to ask any questions they had about using the touch screen tablet during the experiment. Each drive lasted approximately 20 minutes and there were three drives. The entire experiment, with questionnaires and debriefing, took participants 1.5 to 2 hours to complete.

3.2.4 Driving Scenario

Each study drive contained anywhere from 15 to 18 four-way intersections. This study focused on intersections as the majority of pedestrian-vehicle crashes occur at intersections [1, 146]. Each block (from one intersection to the next) corresponded to a distance of approximately 100 meters. At six of these intersections, the simulator was paused to allow participants to answer a set of SAGAT questions on a tablet display. The participants were not informed as to which intersections would be paused. The intersection scenarios varied based on four independent variables described below.

Number and location of a pedestrian (6 levels)

All intersection scenarios included conflicting as well as non-conflicting pedestrians to add a sense of realism to the scenarios. However, SAGAT responses were solely gathered for conflicting pedestrians, as each paused scene deliberately encompassed areas featuring conflicting pedestrians only.

- 1 conflicting pedestrian + n non-conflicting pedestrians (4 levels).
  This scenario included one pedestrian in conflict with the route of the ego car and an additional n non-conflicting pedestrians, where n is 0, 1, or 2. The conflicting pedestrian was positioned in one of four different corner locations (levels) relative to the ego car's location. The non-conflicting pedestrians were positioned randomly but consistently for each scenario. For example, if the non-conflicting pedestrian was placed in the northeast corner for scenario A, it would always be placed in the northeast corner for scenario A (Figure 3.1).

- 2 conflicting pedestrians + n non-conflicting pedestrians (2 levels).
  This scenario included two pedestrians in conflict with the route of the ego car and an additional n non-conflicting pedestrians. The conflicting pedestrians were positioned in two different locations (levels): close to or far from the ego car. The non-conflicting pedestrians were also positioned randomly and consistently within a scenario.

Distance of a pedestrian from the crossing (2 levels: near, far)

This independent variable defines the proximity of the pedestrian to the crossing of the intersection. 'Near' is defined as a conflicting pedestrian located less than 9.5 meters from the crossing. 'Far' indicates a conflicting pedestrian who is observable by the participant but located further than 9.5 meters from the crossing (Figure 3.2).

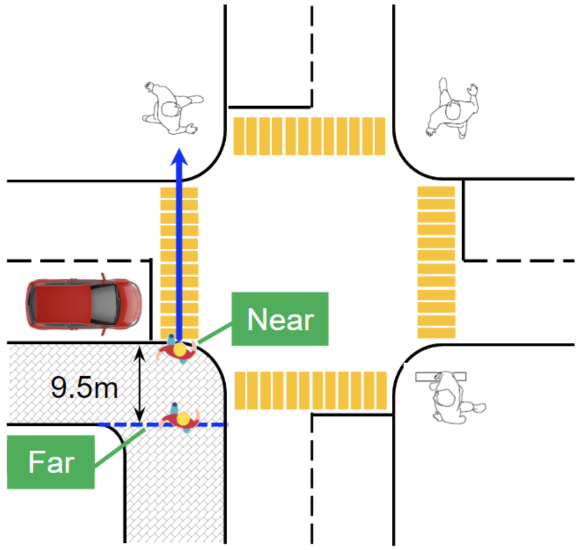


Figure 3.2: Distance of a pedestrian from the crossing

Pedestrian direction (2 levels: clear, unclear)

This independent variable was used to indicate whether the pedestrian's intent to cross was clear or not.

Pedestrian motion (2 levels: walking, standing)

This independent variable defines the action of the pedestrian when the scenario was paused. If the pedestrian was in 'walking' mode, the pedestrian was either crossing the street or approaching the crossing. In 'standing' mode, the conflicting pedestrian was standing at their location.

A combination of 48 different driving scenarios ($6 \times 2 \times 2 \times 2$) was generated for the study. This provided a wide selection of situations for participants to experience.

3.2.5 SAGAT Questionnaire

When the simulator was paused at each driving scenario, the pedestrians were hidden from the driving simulator screens (Figure 3.3).

At each pause, a driver answered questions related to SA levels 1 to 3 using a touch screen tablet. SA is defined as "the perception of environmental elements and events with respect to time or space (level 1), the comprehension of their meaning (level 2), and the projection of their future status (level 3)" [29]. There were five questions in total: two SA level 1 questions, one SA level 2 question, one SA level 3 question, and a confidence level (Figure 3.4). Participants were asked to answer these questions for every pedestrian observed. The questions are shown in Table 3.1.

Table 3.1: SAGAT Questionnaire

	Question	Answer	SA Level
1	Please touch the area where you have seen pedestrians. Touch all that apply and answer the questions accordingly.	1–8	1
2	Which direction was the pedestrian moving toward (if the pedestrian was standing still, please select their facing direction)?	Left, right, same, toward	1
3	Do you think the pedestrian will cross the street so that you will need to stop?	Yes/No	2
4-1	If you made no change in speed, do you think you will collide with the pedestrian?	Yes/No	3
4-2	How confident are you with the response?	0–100	3

The participants also completed questionnaires on satisfaction and demand after each study drive.

3.2.6 Balanced Incomplete Block Design (BIBD)

A BIBD was employed in this study because it was not feasible to include all possible combinations of factors in every block. To control for individual differences based on gender, the number of sessions experienced by each gender was approximately balanced.

The 48 driving scenarios were divided into 8 sessions, with each session consisting of 6 scenarios. Each participant experienced 3 out of the 8 sessions, resulting in a total of 18 scenarios per participant (3 study drives × 6 scenarios). Consequently, each scenario was repeated 21 times for the study, ensuring an adequate sample size for analysis and improving the reliability of the findings.

3.2.7 Measures

Dependent variable

The dependent variable in the study was the participants' responses to the SAGAT questionnaire. There were four SAGAT responses corresponding to SA Level 1 to 3. SA Level 1 had two parts pertaining to the pedestrian's direction and location, SA Level 2 pertained to the pedestrian's intention of crossing, and SA Level 3 pertained to the prediction of collision with the ego car. Each response was recorded as a binary variable: 1 for correct, 0 for incorrect. Confidence levels were also collected for comparisons to other SA studies [127, 39, 145].

There were 1,830 scores collected from study participants. After removing observations with missing driving performance data, 1,808 scores were available for data analysis.

The summary statistics showed that SA Level 2 questions had the lowest correct response rate compared to SA Level 1 and 3 (see Table 3.2 and Figure 3.5).

Table 3.2: All SA responses

Response	SA level 1		SA level 2	SA level 3
	Location	Direction		
Correct	1,198 (66.3%)	1,022 (56.5%)	872 (48.2%)	927 (51.3%)
Wrong	610	786	936	881
Total	1,808	1,808	1,808	1,808

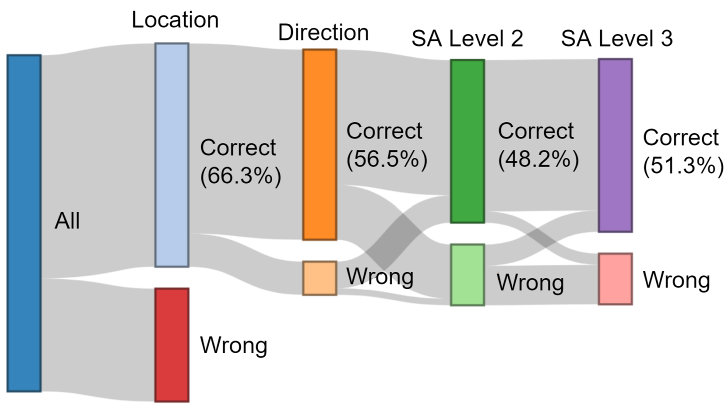


Figure 3.5: Flow chart of SA responses

Figure 3.6 shows the variability in the percentage of correct SA responses. This variability is observed across scenario orders 1 to 6 within each of the 8 driving scenario sessions, with distinct reference to each level of SA questions. As noted earlier, participants were provided breaks following the completion of each session containing 6 driving scenarios. Given this setup, the analysis compared the correct response rates within each session (rather than across different sessions) to gain insights into potential learning effects that might have influenced participants' accuracy over time.

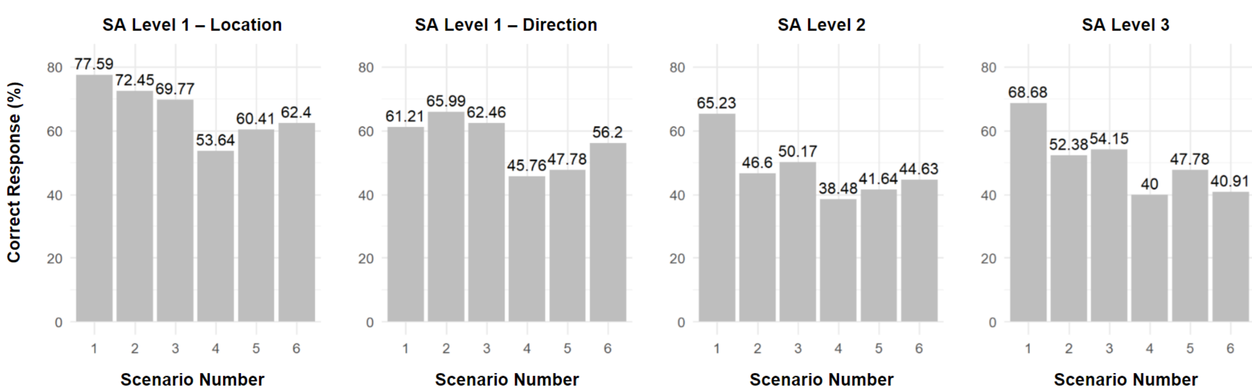


Figure 3.6: Correct SA responses (%) across scenario order 1 to 6

Independent variables

- Driving Performance Measures

  There is currently a lack of consensus on the optimal time window to capture the relationship between driver SA and driving performance measures [46, 45, 98, 136]. The temporal differences that may exist was initially reviewed using a range of summary statistics. These included the minimum, maximum, mean, standard deviation (SD) of the driving performance measures over the specified time windows of 1, 3, and 5 seconds. Deviations in the driving performance measures were computed as the difference between each summary statistics at intervals of 0.1 seconds.

  Speed of the ego car.

  This is a continuous measure, ranged from 0 to 25 mph. It has been shown that drivers display greater speed variability and throttle control while in a distracted driving situation [51, 103, 138, 10]. As speed is a continuous measure, the mean, min, max, and standard deviation were reviewed for modeling as they have been considered in other studies [82].

  Acceleration pedal input.

  This is a continuous measure which ranged from 0 to 1, where 0 indicates that the gas pedal has been released completely and 1 indicates that the gas pedal has been pressed all the way. Distracted drivers have been shown to apply the pedals differently when compared to non-distracted drivers [60, 124]. Studies also show lower standard deviation of accelerator pedal when drivers were distracted [51, 139].

  Brake pedal input.

  This is a continuous measure which ranged from 0 to 1, where 0 indicates that the brake pedal has been released completely and 1 indicates that the brake pedal has been pressed all the way. It has been shown that when drivers are distracted, they show more occurrences of hard braking [54, 53, 77].

Steering wheel input.

This is a continuous measure which ranged from -1 to 1, where -1 indicates that the wheel is turned to the left most position and 1 indicates that the wheel is turned to the right most position. Measures of steering wheel control have included the standard deviation of steering wheel angle, steering wheel reversal rate, steering wheel action rate, and steering entropy. When driving without any distraction source, drivers make a number of small corrective steering wheel movements to maintain lateral position while in distracted driving drivers often make a number of large and abrupt steering wheel movements to correct driving errors [108, 13, 87, 105, 115].

Lane deviation.

This measure refers to the lateral position of the vehicle on the road in relation to the center of the lane in which the vehicle is traveling. Relationships between the lane deviation of the vehicle and driver distraction have been consistently observed in past studies [47, 57, 15, 72].

- Pedestrian Interaction

  Direction (5 levels).

  The levels of the variable include clear-same (walking or standing in the same direction as the ego car), clear-toward (walking or standing toward the ego car), clear-left, clear-right, unclear (all other directions). The variable was used to indicate whether the pedestrian's intent to cross was evident or unclear when generating driving scenarios. Post hoc, the analyst recognized that the "clear" intention needs to be examined in the context of the intended travel direction: same, toward, left, and right.

  Motion (2 levels).

  The levels of the variable include standing and walking.

  Block location (2 levels).

  The levels of the variable include current and next.

Corner location (4 levels).

The levels of the variable include close-left, close-right, far-left and far-right (Figure 3.7).

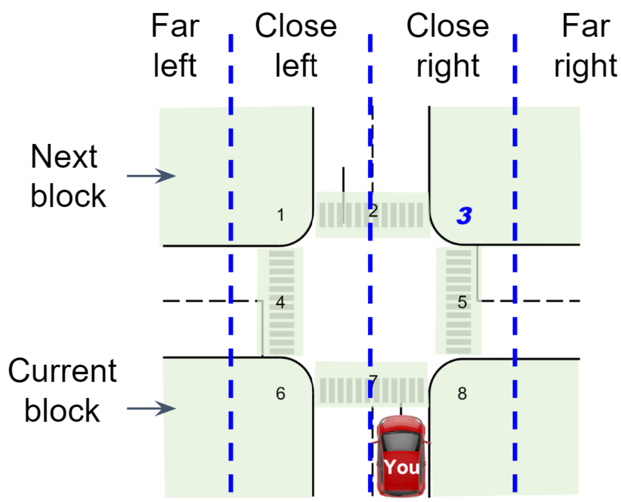


Figure 3.7: Block and corner location of pedestrians

Total number of pedestrians at the intersection (3 levels: 1, 2, 3).

3.2.8 Data Analysis

Out of the 1,830 SA scores collected from the 56 drivers, 1,808 scores were used in the forthcoming analysis after removing rows with missing data.

The `glmer` function from the `lme4` package in R was used to build the models [8]. Mixed effects logistic regression was applied to model the binary response variables, which indicated whether the driver answered correctly or not for varying levels of SA questions. There were separate models for each SA level: pedestrian location (SA level 1), direction (SA level 1), the intention of crossing (SA level 2), and collision prediction (SA level 3). The different time windows (1, 3, and 5 seconds) were also modeled separately, and assessed using the

Akaike Information Criterion (AIC). In total, there were 12 models (4 SA levels x 3 time windows).

Recursive Feature Elimination (RFE) was employed for feature selection to account for multicollinearity among the driving performance measures. The statistical software package `R v4.2.1` was used for all data processing and analysis. The R package `caret` was used for feature selection, employing the `rfeControl` function. The variables were rescaled to ensure similar scales for analysis [67].

3.3 Results

3.3.1 Pedestrian Location and Direction (SA level 1) – Table 3.3, 3.4

Regardless of the time windows, the models consistently showed that driving performance measures and pedestrian interaction factors (direction, speed, block, corner location, and the total number of pedestrians) have a significant impact on the likelihood of drivers accurately identifying the pedestrian location and direction.

For the following results explanation, odds ratios (OR) were employed as a measure of association. OR greater than 1 indicates that the event is more likely to occur as the predictor increases, signifying a positive relationship between the predictor and the event. On the other hand, OR less than 1 suggests that the event is less likely to occur as the predictor increases, indicating a negative relationship between the predictor and the event.

Driving Performance Measures

Drivers were more likely to accurately identify the location of a pedestrian with:

- lower maximum speed values and smaller maximum values for accelerator deviation within a 1-second window (OR = 0.63, 0.25).

- lower maximum speed values and higher minimum values for accelerator deviation and higher maximum brake input within a 3-second window (OR = 0.73, 2.83, 5.62).

- lower average speed values and smaller maximum values for accelerator deviation within a 5-second window (OR = 0.67, 0.36).

Drivers were more likely to accurately identify the direction of a pedestrian with:

- lower minimum speed values and smaller average values for steering wheel deviation within a 1-second window (OR = 0.76, 0.80).

- higher maximum brake input and smaller average values for steering wheel deviation within a 3-second window (OR = 8.41, 0.67).

- lower minimum speed values and smaller minimum values and greater average values for brake deviation within a 5-second window (OR = 0.86, 0.96, 3.90).

Pedestrian Interaction

Drivers were less likely to accurately identify the location and direction of a pedestrian when the pedestrian is walking or standing toward the ego car, compared to when the pedestrian is moving to the left, right, or in a unclear direction.

Drivers are more likely to accurately identify the location and direction of a pedestrian when the pedestrian is:

- standing, compared to walking.

- on the next block, as opposed to on the current block.

- on the close right side of the corner, as opposed to on the close or far left side.

The presence of more pedestrians in the intersection decreases the likelihood of drivers accurately identifying the location and direction of a pedestrian.

3.3.2 Pedestrian Intention of Crossing (SA level 2) – Table 3.5

The model using a 1-second time window did not reveal any significant associations between driving performance measures and the likelihood of drivers accurately identifying the pedestrian's intention of crossing. Only pedestrian interaction factors were shown to be significant in this model.

On the other hand, the models with 3-second and 5-second time windows revealed that both driving performance measures and pedestrian interaction factors have a significant impact on the likelihood of drivers accurately identifying the pedestrian's intention of crossing.

Driving Performance Measures

Drivers are more likely to accurately identify the pedestrian's intention of crossing with:

- smaller average values for steering wheel deviation and a greater maximum distance of the ego car from the lane center within a 3-second window (OR = 0.67, 1.03).

- a greater maximum distance of the ego car from the lane center within a 5-second window (OR = 1.02).

Pedestrian Interaction

Drivers are less likely to accurately identify the pedestrian's intention of crossing when the pedestrian is walking or standing in the same direction as the ego car, compared to when the pedestrian is moving to the left, right, toward, or in an unclear direction.

Drivers are more likely to accurately identify the pedestrian's intention of crossing when the pedestrian is:

- standing, compared to walking.

- on the current block, as opposed to on the next block.

- on the close right side of the corner, as opposed to on the close left, far left, or far right side.

3.3.3 Pedestrian Collision Prediction (SA level 3) – Table 3.6

The models with different time windows (1-second, 3-second, and 5-second) consistently show that both driving performance measures and pedestrian interaction factors have a significant impact on the likelihood of drivers accurately predicting whether they will collide with the pedestrian or not.

Driving Performance Measures

Drivers are more likely to accurately predict the occurrence of pedestrian collision with:

- smaller standard deviation values for the accelerator, smaller maximum values for the steering wheel, and a greater maximum distance of the ego car from the lane center within a 1-second window (OR = 0.22, 0.81, 1.02).

- smaller minimum values for the steering wheel, smaller average values for steering wheel deviation, and a greater maximum distance of the ego car from the lane center within a 3-second window (OR = 0.79, 0.51, 1.02).

- lower average speed values, smaller maximum values for the steering wheel, and greater maximum distance from the lane center within a 5-second window (OR = 0.85, 0.78, 1.03).

Pedestrian Interaction

Drivers are less likely to accurately predict the occurrence of pedestrian collision when the pedestrian is walking or standing in the same direction as the ego car, compared to when the pedestrian is moving to the left, right, toward, or in an unclear direction.

Drivers are more likely to accurately predict the occurrence of pedestrian collision when the pedestrian is:

- standing, compared to walking.
- on the current block, as opposed to on the next block.
- on the close right side of the corner, as opposed to on the close left, far left, or far right side.

The presence of more pedestrians in the intersection decreases the likelihood of drivers accurately predicting the occurrence of pedestrian collision.

3.4 Discussion

This study provides insight on factors associated with the three levels of driver SA: perception, comprehension, and projection. The factors included driving performance measures and pedestrian interaction. The analysis of driving performance measures from a driving simulator study highlighted the usefulness of different time windows in predicting driver SA. This study proposed an alternative model that can be used in cases where the pedestrian visible time window varies and is difficult to track accurately, which aligns with real-world driving scenarios.

This study showed that a strong correlation exist among driving performance measures such as the maximum lane deviation and steering wheel input. Exploring the applicability of these measures, particularly in scenarios with limited data or specific time windows, emerges as a compelling avenue for future research.

The findings showed that a single standard time window is not as useful for predicting driver SA. More specifically, tailoring the time windows to different levels of driver SA can provide better prediction performance. In this study, a model using a 3-second window outperformed models using 1-second and 5-second windows when predicting driver SA regarding pedestrian location and crossing intent. That said, the 1-second model was useful when the focus was on pedestrian direction. For forecasting driver SA concerning pedestrian collisions, the 5-second model exhibited superior performance.

For SA Level 1 (perception), no differences in correct responses were observed between the *same* and *toward* pedestrian direction. This can be explained by the ease in correctly perceiving the location and direction of pedestrians in our study, regardless of the pedestrian's direction [122].

For SA Levels 2 and 3 (comprehension and projection) and regardless of the time window, drivers were more likely to answer correctly when the pedestrian was facing the ego car when compared to traveling in same direction. As expected, it is easier to see and predict the path of a pedestrian moving toward the driver compared to moving away from the driver.

For SA Level 1 response, drivers were more likely to answer correctly when the pedestrian location was on the next block compared to the current block. However, for SA Level 2 and 3, the opposite was observed. This can be attributed to the fact that if the pedestrian is in the current block, they are close enough to provide sufficient information about their intention to cross, making it easier for drivers to predict. On the other hand, if the pedestrian is on the next block, there is still a chance that they may attempt to cross before the driver reaches the intersection, introducing some uncertainty [5, 134].

In this study, distinct models were developed to address varying levels of driver SA. A model predicting an aggregated SA response across all SA Levels was not developed within the study's scope. This challenge arises due to the non-linear relationship between driver SA Levels, despite their ascending order from level 1 to 3. Moreover, the relative importance of these levels could vary based on the context and driving scenario. Given these complexities, further research is needed to devise a model capable of predicting an aggregated SA response that encompasses all SA Levels.

One limitation of the study is the potential misunderstanding of the question structure. Although extensive pilot testing was conducted, there is still a possibility that drivers may misinterpret the pedestrians' crossing intent. Different drivers may have varying interpretations of the same scenario, which are based on subjective reasons that were not captured.

In future studies, it would be important to capture their reasoning as well as exploring the methods to minimize subject interpretation biases. Enhancing the consistency and accuracy of measuring pedestrian intention would contribute to more reliable and robust findings in driver SA predictive models.

We acknowledge that the driving simulator environment, while controlled, may not capture all complexities of real-world driving scenarios with pedestrian interactions. The driving scenarios in this study were limited to intersections and the findings may not generalize to other contexts like mid-block pedestrian crossings.

The study's focus on SA during active driving tasks limited its scope to SAE Level 2 automation and below. Considering the prevalence of SAE Level 2 on roads, the study recognizes the continuing relevance of its findings, even as higher levels of automation emerge. It is important to recognize that over the next several years not all mobility services will provide high or full autonomy (SAE Level 4 or 5 [19]) but several services may remain partially automated.

In terms of future exploration, it would be useful to account for more diverse pedestrian speeds, including variations between standing and walking. Further investigation should also consider scenarios with more complex pedestrian crossing and traffic flow, mirroring real-world situations and the challenges for better situational awareness.

As with many other study environments, this setting was constrained by a limited number of participants, necessitating the use of available data and AIC to evaluate the model's adequacy. While the application of AIC for model comparison is appropriate, the authors acknowledge that, in the absence of a power analysis, the sample size may have been insufficient to detect an effect. Future research would benefit from employing larger sample sizes.

3.5 Conclusion

This study aimed to investigate the key factors in predicting driver SA, encompassing driving performance measures and pedestrian interaction factors. By employing the SAGAT method to assess drivers' perception, comprehension, and projection of pedestrian behavior across different SA Levels, ranging from level 1 to 3, we identified significant contributions from both driving performance measures and pedestrian interaction factors in predicting driver SA. The result suggests that different time windows are useful for predicting various levels of driver SA responses, with no universally applicable best time window. Comparing 12 different models, this study evaluated their goodness of fit and significant factors. For example, the 3-second model excelled in predicting correct pedestrian location and crossing intention, whereas the 1-second model outperformed others in predicting pedestrian direction. These findings establish a foundation for future studies on employing SA in complex driving environments and provide insights for building predictive models of driver SA in various situations.

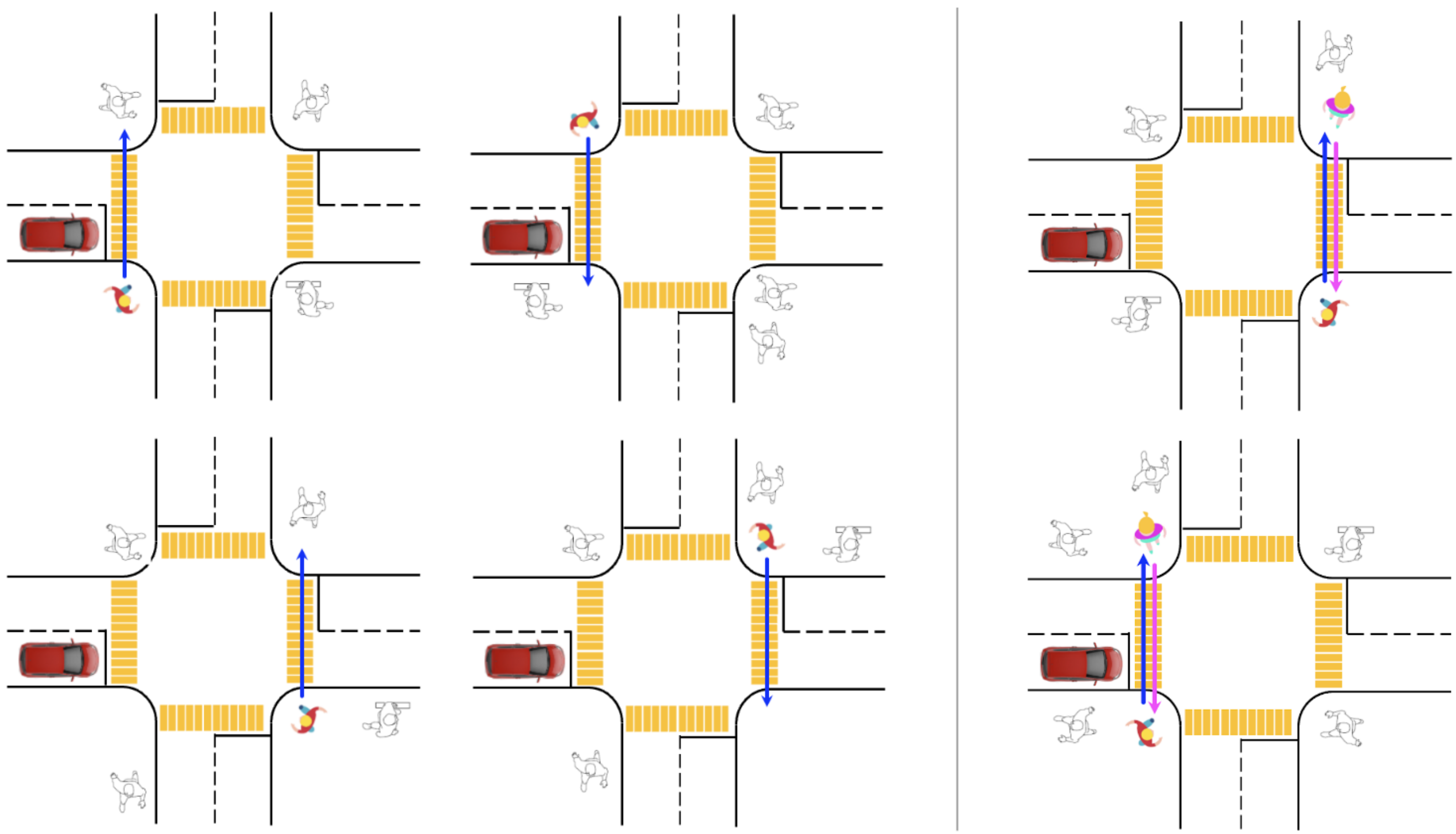


Figure 3.1: 1 conflicting + n non-conflicting pedestrians (left) and 2 conflicting + n non-conflicting pedestrians (right)

Figure 3.3: Actual scene (top) and Empty scene at a pause with a pedestrian hidden (bottom)

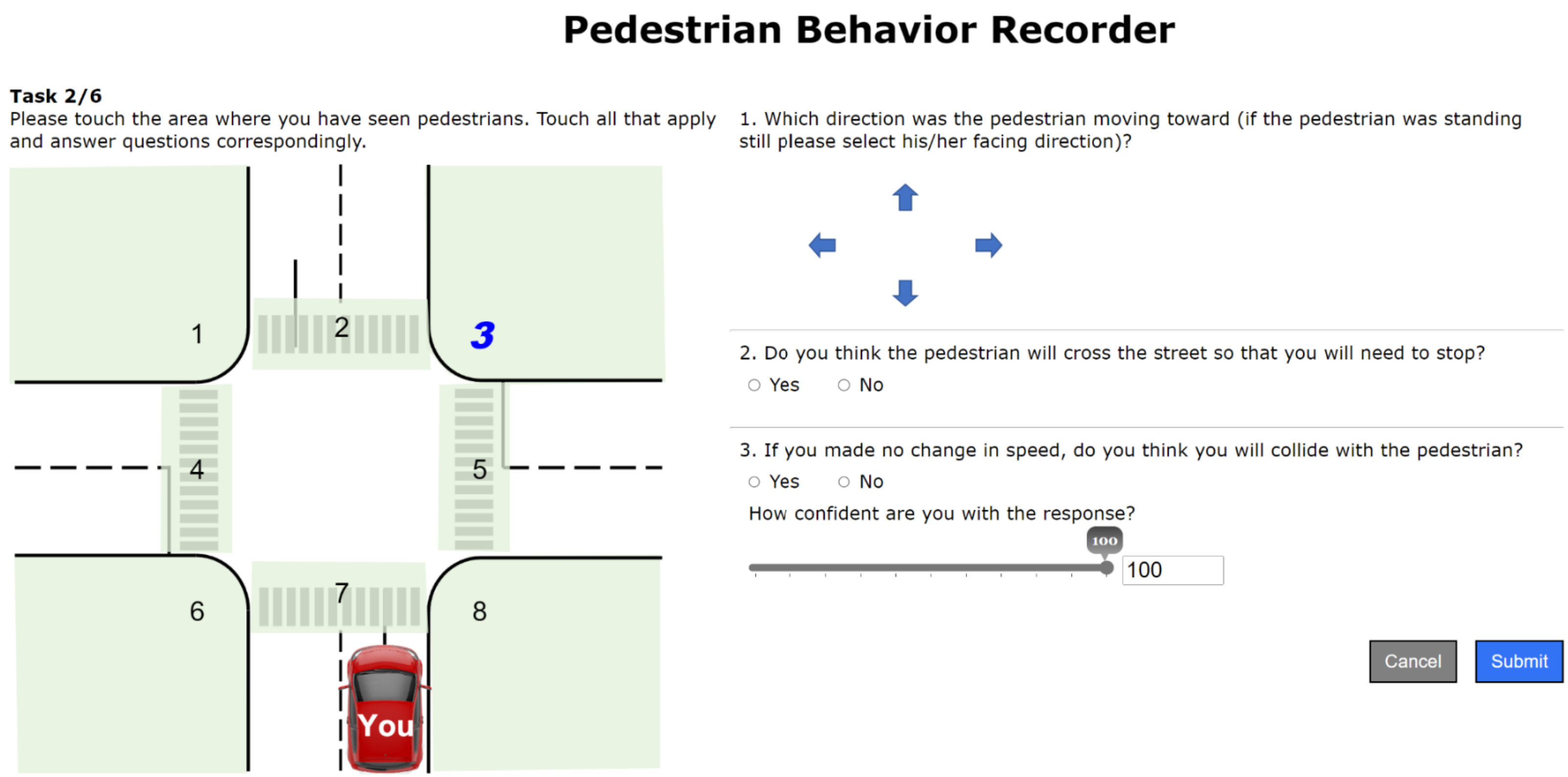


Figure 3.4: Example of a response when area 3 is selected in SAGAT Questionnaire.

Table 3.3: Likelihood of selecting correct pedestrian location (yes, no) - SA level 1

	1 sec			3 sec			5 sec		
	Estimate		OR	Estimate		OR	Estimate		OR
Intercept	2.15	***	8.58	1.71	***	5.54	1.95	***	7.02
Driving Performance Measures									
Speed Max	-0.47	***	0.63	-0.32	**	0.73			
Speed Mean							-0.40	***	0.67
Accelerator Deviation Max	-1.39	*	0.25				-1.01	*	0.36
Accelerator Deviation Min				1.04	**	2.83			
Brake Max				1.73	*	5.62			
Pedestrian Interaction									
Direction									
Clear-toward (reference)									
Clear-same	0.21		1.23	0.21		1.24	0.21		1.24
Clear-left	1.12	***	3.05	1.14	***	3.12	1.12	***	3.07
Clear-right	1.07	***	2.91	1.11	***	3.04	1.10	***	3.00
Unclear	0.85	***	2.34	0.87	***	2.39	0.89	***	2.43
Motion									
Standing (reference)									
Walking	-0.79	***	0.45	-0.79	***	0.45	-0.78	***	0.46
Block									
Current (reference)									
Next	0.22		1.24	0.25	*	1.28	0.24	*	1.27
Corner									
Close-left (reference)									
Close-right	0.33	*	1.39	0.35	*	1.42	0.37	*	1.44
Far-left	-1.48	***	0.23	-1.48	***	0.23	-1.45	***	0.23
Far-right	0.12		1.13	0.11		1.12	0.14		1.15
Total # of Ped	-0.28	**	0.75	-0.26	**	0.77	-0.28	**	0.75
AIC	1958.77			1951.44			1957.04		
Log Likelihood Initial	-1125.07			-1125.07			-1125.07		
Log Likelihood Convergence	-965.38			-960.72			-964.52		
SD of Random Effect	0.51			0.51			0.51		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3.4: Likelihood of selecting correct pedestrian direction (yes, no) - SA level 1

	1 sec			3 sec			5 sec		
	Estimate		OR	Estimate		OR	Estimate		OR
Intercept	1.64	***	5.17	0.93	**	2.55	1.31	***	3.72
Driving Performance Measures									
Speed Min	-0.27	***	0.76				-0.15		0.86
Brake Max				2.13	***	8.41			
Brake Deviation Min							-0.05	*	0.96
Brake Deviation Mean							1.36	***	3.90
Steering Deviation Mean	-0.23	*	0.80	-0.41	*	0.67			
Pedestrian Interaction									
Direction									
Clear-toward (reference)									
Clear-same	0.30		1.35	0.44		1.56	0.33		1.40
Clear-left	1.25	***	3.47	1.31	***	3.71	1.27	***	3.58
Clear-right	1.14	***	3.11	1.21	***	3.34	1.14	***	3.13
Unclear	1.01	***	2.73	1.05	***	2.87	1.03	***	2.80
Motion									
Standing (reference)									
Walking	-1.14	***	0.32	-1.15	***	0.32	-1.15	***	0.32
Block									
Current (reference)									
Next	0.49	***	1.64	0.61	***	1.84	0.52	***	1.69
Corner									
Close-left (reference)									
Close-right	0.30		1.34	0.28		1.32	0.28		1.32
Far-left	-1.41	***	0.24	-1.44	***	0.24	-1.44	***	0.24
Far-right	0.14		1.15	0.11		1.11	0.09		1.10
Total # of Ped	-0.59	***	0.55	-0.57	***	0.56	-0.60	***	0.55
AIC	2007.63			2021.44			2011.01		
Log Likelihood Initial	-1208.89			-1208.89			-1208.89		
Log Likelihood Convergence	-988.81			-996.72			-990.5		
SD of Random Effect	0.37			0.38			0.40		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3.5: Likelihood of selecting correct pedestrian intention of crossing (y,n) - SA level 2

	1 sec			3 sec			5 sec		
	Estimate		OR	Estimate		OR	Estimate		OR
Intercept	0.23		1.25	-0.08		0.92	-0.05		0.95
Driving Performance Measures									
Steering Deviation Mean				-0.40	*	0.67			
Lane Deviation Max				0.03	*	1.03	0.02	*	1.02
Pedestrian Interaction									
Direction									
Clear-toward (reference)									
Clear-same	-1.03	**	0.36	-0.94	**	0.39	-0.96	**	0.38
Clear-left	0.58	**	1.78	0.63	***	1.88	0.61	***	1.85
Clear-right	0.91	***	2.50	0.98	***	2.67	0.93	***	2.54
Unclear	0.83	***	2.29	0.88	***	2.42	0.87	***	2.39
Motion									
Standing (reference)									
Walking	-1.16	***	0.31	-1.13	***	0.32	-1.15	***	0.32
Block									
Current (reference)									
Next	-0.39	***	0.67	-0.42	***	0.66	-0.41	***	0.67
Corner									
Close-left (reference)									
Close-right	0.43	**	1.54	0.46	**	1.59	0.43	**	1.53
Far-left	-0.54	***	0.58	-0.55	***	0.58	-0.55	***	0.58
Far-right	0.39	*	1.48	0.41	*	1.51	0.39	*	1.48
Total # of Ped	-0.04		0.96	-0.00		1.00	-0.01		0.99
AIC	2161.14			2154.36			2157.22		
Log Likelihood Initial	-1212.57			-1212.57			-1212.57		
Log Likelihood Convergence	-1068.57			-1063.18			-1065.61		
SD of Random Effect	0.47			0.47			0.49		

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Table 3.6: Likelihood of selecting correct pedestrian collision prediction (y,n) - SA level 3

	1 sec		3 sec		5 sec	
	Estimate	OR	Estimate	OR	Estimate	OR
Intercept	0.60	1.82	0.42	1.52	0.82 *	2.28
Driving Performance Measures						
Speed Mean					-0.16 *	0.85
Accelerator SD	-1.50 *	0.22				
Steering Max	-0.21 **	0.81			-0.25 ***	0.78
Steering Min			-0.23 *	0.79		
Steering Deviation Mean			-0.67 **	0.51		
Lane Deviation Max	0.02 *	1.02	0.02 *	1.02	0.03 **	1.03
Pedestrian Interaction						
Direction						
Clear-toward (reference)						
Clear-same	-1.50 ***	0.22	-1.51 ***	0.22	-1.57 ***	0.21
Clear-left	0.82 ***	2.27	0.79 ***	2.21	0.80 ***	2.23
Clear-right	1.15 ***	3.15	1.13 ***	3.10	1.13 ***	3.09
Unclear	0.70 ***	2.02	0.66 ***	1.94	0.71 ***	2.03
Motion						
Standing (reference)						
Walking	-1.21 ***	0.30	-1.20 ***	0.30	-1.21 ***	0.30
Block						
Current (reference)						
Next	-0.44 ***	0.64	-0.44 ***	0.65	-0.53 ***	0.59
Corner						
Close-left (reference)						
Close-right	0.49 **	1.63	0.48 **	1.61	0.52 **	1.68
Far-left	-0.54 ***	0.58	-0.53 ***	0.59	-0.47 **	0.63
Far-right	0.89 ***	2.44	0.91 ***	2.49	0.92 ***	2.52
Total # of Ped	-0.19 *	0.83	-0.18	0.84	-0.19 *	0.83
AIC	2084.47		2082.61		2080.45	
Log Likelihood Initial	-1207.14		-1207.14		-1207.14	
Log Likelihood Convergence	-1027.24		-1026.31		-1025.22	
SD of Random Effect	0.55		0.55		0.54	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Chapter 4

DEVELOPMENT OF SA PREDICTIVE MODELS

4.1 Introduction

This chapter focuses on the development of predictive models of driver SA. The models' original response is binary, indicating whether the driver SA is classified as low or high across SA levels 1 to 3 (coded as 0 for low and 1 for high). The determination of SA level is based on the correctness of the driver's response, where a correct answer signifies high SA, and an incorrect answer signifies low SA. The ground truth data for SA was obtained through a driving simulator study utilizing Situation Awareness Global Assessment Technique (SAGAT), as detailed in Chapter 3. This chapter presents the development of comprehensive predictive models with response variables categorized into low, mid, and high SA, to cover all levels of SA.

The features employed in the model development encompass a diverse set, including driving performance measures, pedestrian interactions (direction, location, speed, etc.), driver eye-gaze metrics (gaze distance from objects, gaze velocity, pupil size, etc.), and demographic information (age, years of driving experience, etc.). The data set utilized for model training, validation and evaluation is derived from the data collected in the two preceding studies.

Certain features related to driver eye-gaze required cleaning and additional processing to extract pertinent information, contributing to the augmentation of features employed in the model development process. Given the heterogeneous nature of the features, comprising both continuous and categorical variables, a tailored approach was adopted to account for these differences in the construction of various models.

4.2 Related Work

Studies have developed SA predictive models using various features. [137, 144] used Eye-gaze features. [99, 136, 135] used environmental features such as roadway type and visual complexity of the road. Limited studies used vehicle parameters for SA prediction [100, 143].

Research have demonstrated the feasibility of model explanation method as a feature selection method across fields from biomedical engineering to US housing. [83] demonstrated that SHAP is superior to other common feature selection mechanisms compared to ANOVA and RFE. [117] showed that LIME outperformed the other feature selection methods in most cases and could achieve a statistically significantly better classification accuracy than that of the benchmark methods. However, there is a limited research on using mixed use of traditional methods and explainabiltiy method (SHAP) for feature selection in developing models of driver SA.

4.3 Dataset

We used the dataset collected in [100] with 56 participants with an equal distribution between those who self-identified as females and non-females. The participants ranged in age from 18 to 52, with an average age of 26. The study consisted of three parts: 1) a pre-experiment questionnaire, 2) a practice drive, and 3) the main study, which included three study drives and a post-experiment questionnaire.

The pre-experiment questionnaire included demographic information (age, gender), driving experience (years of experience, miles driven/month), and the Driver Behavior Questionnaire (DBQ) [107].

To familiarize participants with the driving simulator, a 5-minute practice drive was conducted under road conditions similar to those in the main study, including intersections with pedestrians.

During the main study, each participant operated the steering wheel and pedals in the driving simulator. Their task was to drive through intersections that were representative of a US urban downtown area, following the road without making any left or right turns. The simulator would pause randomly during each scenario so that prediction of the pause location would be minimized. At each pause, drivers would pick up the touch screen tablet and answer SAGAT questions. Each drive lasted approximately 20 minutes and there were three drives. Each study drive contained anywhere from 15 to 18 four-way intersections.

The responses from SAGAT questions were used to create the ground truth of SA of the participants. At each pause, there were five questions: two SA Level 1 questions, one SA Level 2 question, one SA Level 3 question, and a confidence level. Each question was constructed based on the examples of SAGAT queries on three levels of SA presented by [33, 35, 36]. Participants were asked to answer these questions for every pedestrian observed. Using the responses, we modeled the SA prediction as a classification problem described in the next section. There were 1,830 scores collected from study participants for each of the SAGAT question.

Tobii Pro Glasses 3 eye tracking device[1] was used to record participants' eye movements. It recorded participants' eye movements at a sampling rate of 100 Hz, and was located 48 inches in front of three 42-inch plasma displays, each with a resolution of 1,280 × 720.

4.4 Method

4.4.1 Data Processing

Capping Outliers: For each numerical column, the lower and upper bounds were determined using the 5th and 95th percentiles, respectively. Values below the 5th percentile were replaced with the value at the 5th percentile, and values above the 95th percentile were replaced with the value at the 95th percentile. This step ensures that extreme values do not adversely affect subsequent analyses.

For columns containing missing values, the median was computed and used to fill in the missing entries.

[1] https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3

Target Encoding: Categorical features were encoded based on the mean target values of each category, enhancing the model's ability to learn from categorical data. This encoding was applied only to the training data to prevent data leakage.

Normalization: Continuous features underwent Min-Max scaling to standardize their ranges, ensuring that all features contribute equally to the model's learning process.

Balancing the Dataset: The Synthetic Minority Over-sampling Technique (SMOTE) was employed to address class imbalance. This technique synthesized new examples of the minority class, allowing the model to learn more effectively from all classes.

4.4.2 Hyperparameter Optimization

To identify optimal model parameters, a parameter grid was defined, focusing on key hyperparameters of the XGBoost model: the number of estimators, max depth of tree, and learning rate. The RandomizedSearchCV was employed to conduct hyperparameter optimization over multiple iterations improving model accuracy.

4.4.3 Cross-Validation

A nested cross-validation framework was used to ensure reliable performance metrics. To mitigate the risk of data leakage, we ensured that data processing adhered to the nested loop structure. The outer loop, consisting of 10 folds, was utilized for model evaluation, while the inner loop also comprised 10 folds for hyperparameter tuning. This dual-layered approach mitigates the risk of overfitting and provides a robust assessment of the model's performance.

4.4.4 Response variable class

There were total of four binary SA responses for each question initially. In effort to develop a predictive model that encompass three levels of SA, the response variable was defined as three different classes, SA low, medium, high based on the responses. The classification was done to ensure the data balance.

4.5 Features

The entire list of features used in model development are listed in Table 4.1, 4.2, 4.3, 4.4, 4.5.

Table 4.1: Features used in predicting SA – Demographics

	Measures	Explanation	Unit	Range
	Demographics			
1	age	Age of the drivers	year	18 - 40
2	gender	Female = 0 ; Male = 1		
3	yearDriving	Years of driving experience	year	1 - 25
4	dFrequency	Driving frequency in last 12 months: at least once daily = 1; at least once weekly = 2; less than once weekly = 3		
5	nViolation	Number of moving violations the driver had in the last 3 years		0 - 5
6	nCrash	Number of vehicular crashes the driver had in the last 3 years		0 - 5
7	confidence	Likert scale on how confident the driver was about their answer: from not confident all = 0 to very confident = 100		

Eye-tracking techniques hold significant potential for estimating driver awareness of road hazards, offering the advantage of real-time application without interrupting the ongoing driving task [65, 66, 91]. The eye-mind hypothesis [61, 111] emphasizes the close relationship between eye gaze patterns and cognitive engagement. Consequently, driver eye-gaze related features are widely incorporated into predictive models for driver SA.

In a study by [144], eye-tracking data was employed to predict driver SA in real-time during takeover transitions in conditionally automated driving. A LightGBM tree ensemble machine learning model was used, revealing that increased fixation numbers on the rear-view mirror and longer fixation durations were associated with better SA. Conversely, a higher fixation on the road was linked to lower SA, possibly indicating a need for participants to examine mirrors for an overall understanding of driving scenarios.

In the work by [65, 66], correlation and regression analyses between indirect (eye-tracking) and direct measures of driver awareness identified variables capturing both spatial and temporal aspects of eye-glance behavior. Gaze distance variables reflected spatial characteristics by measuring angular distances of road hazards from the driver's gaze position, while fixation time variables captured temporal aspects of eye-glance behavior.

Despite the promising aspects of eye-gaze data, it is acknowledged that this information alone is insufficient for accurate SA prediction, necessitating integration with other factors for a comprehensive model.

In Study 2, driver eye-gaze data was collected using a Tobii eye tracker, capturing gaze position in x-y coordinates, pupil sizes for the left and right eyes, gaze velocity, and the fixation status (whether the gaze is fixed or not). Saccades and fixations were differentiated based on a gaze velocity threshold of 30 degrees/second. This eye-gaze data was subsequently synchronized with the recordings of three screens in the driving simulator, enabling the extraction of gaze points in the driving scene in x-y coordinates for further analysis (Fig 4.1).

4.6 Models

The models considered for the development of the predictive framework include, but are not limited to, the following:

Figure 4.1: Tobii eye tracker was used to collect drvier eye-gaze data. The top image is eye tracker recording from driver and the bottom image is the three screens of driving simulator.

1. Logistic Regression: Logistic Regression is a linear model specifically designed for binary classification problems. It estimates the probability that an instance belongs to a particular class, transforming the output using the logistic function to produce a probability score between 0 and 1. This algorithm is well-suited for scenarios where the assumed relationship between features and the log-odds of the response variable is linear.

2. Linear Support Vector Machine (SVM): Support Vector Machine (SVM) is a versatile supervised learning algorithm applicable to both classification and regression tasks. The linear SVM seeks to identify the optimal hyperplane that maximizes the margin between different classes in the feature space. It excels in high-dimensional spaces and demonstrates robustness to outliers.

3. Quadratic SVM: Quadratic SVM extends the linear SVM by incorporating quadratic

features, enabling the modeling of non-linear decision boundaries. This is particularly advantageous when the relationship between features and the response variable is intricate and cannot be adequately captured by a linear model.

4. Gaussian SVM: Gaussian SVM, or Radial Basis Function (RBF) SVM, is an extension of SVM utilizing a non-linear kernel (Gaussian kernel). This kernel maps input features into a higher-dimensional space, allowing the algorithm to capture complex relationships between features and the response variable. It is well-suited for non-linear classification tasks.

5. Decision Tree (Medium Tree): A Decision Tree is a tree-like model where each node represents a decision based on a feature, leading to a final prediction at the leaf nodes. "Medium Tree" refers to a decision tree of moderate depth, striking a balance between capturing complexity and avoiding over-fitting.

6. Random Forest: Random Forest is an ensemble learning method that constructs multiple decision trees and aggregates their predictions. It introduces randomness in the tree-building process, enhancing robustness and reducing susceptibility to over-fitting compared to a single decision tree.

7. Adaptive Boosting (AdaBoost): AdaBoost iteratively trains the weak classifier on the training data set with each successive classifier giving more weight to the data points that are misclassified. The final AdaBoost model is decided by combining all the weak classifier that has been used for training with the weight given to the models according to their accuracies. The weak model which has the highest accuracy is given the highest weight while the model which has the lowest accuracy is given a lower weight.

8. Gradient Boosting: Gradient Boosting is a powerful boosting algorithm that combines several weak learners into strong learners, in which each new model is trained to minimize the loss function such as mean squared error or cross-entropy of the previous

model using gradient descent. In contrast to AdaBoost, the weights of the training instances are not tweaked, instead, each predictor is trained using the residual errors of the predecessor as labels.

Table 4.2: Features used in predicting SA – Pedestrian Interaction

	Measures	Explanation	Unit	Range
	Pedestrian Interaction			
8	direction	Whether the pedestrian's intent to cross was clear or not (5 levels: clear-same, clear-toward, clear-left, clear- right, unclear)		
9	motion	Action of the pedestrian when the scenario was paused (2 levels: walking, standing)		
10	block	The levels of the variable include current and next.		
11	corner	The levels of the variable include close-left, close-right, far-left and far-right		
12	nPed	Total number of pedestrians at the intersection (3 levels: 1, 2, 3)		1 - 3
13	dCenterEu	Euclidean distance of a pedestrian from the crossing center		
14	dCenterX	Lateral distance of a pedestrian from the crossing center		
15	dCenterY	Longitudinal distance of a pedestrian from the crossing center		
16	dEgoEu	Euclidean distance of the ego car from the crossing center		
17	dEgoX	Lateral distance of the ego car from the crossing center		
18	dEgoY	Longitudinal distance of the ego car from the crossing center		

Table 4.3: Features used in predicting SA – Driving Performance

	Measures	Explanation	Unit	Range
	Driving Performance Measures			
19	sMax	Maximum speed of the ego car	mph	0 - 25
20	sMean	Average speed of the ego car		
21	sMin	Minimum speed of the ego car		
22	aStd	Standard deviation of acceleration pedal input		0 - 1
23	aDevMean	Average value of deviation of acceleration pedal input (per 0.1 sec)		
24	aDevMin	Minimum value of deviation of acceleration pedal input (per 0.1 sec)		
25	bMax	Maximum brake pedal input		
26	bDevMean	Average value of deviation of brake pedal input (per 0.1 sec)		
27	bDevMin	Minimum value of deviation of brake pedal input (per 0.1 sec)		
28	strMax	Maximum value of steering wheel input		0 - 1
29	strMin	Minimum value of steering wheel input		-1 - 0
30	strDevMean	Average value of deviation of steering wheel input (per 0.1 sec)		
31	lDevMax	Minimum deviation of the lateral position of the vehicle on the road in relation to the center of the lane in which the vehicle is traveling (per 0.1 sec)		

Table 4.4: Features used in predicting SA – Eye-gaze

	Measures	Explanation	Unit	Range
	Eye-gaze			
32	fMean	Average fixation duration of the driver		
33	fStd	Standard deviation of fixation duration of the driver		
34	fMax	Maximum value of fixation duration of the driver		
36	road	Number of fixations of the driver on the road		
36	sky	Number of fixations of the driver on the sky		
37	mirror	Number of fixations of the driver on the mirror		
38	numF	Number of gaze fixations of the driver during certain period of time by the pause		

Table 4.5: Features used in predicting SA – Eye-gaze (continued)

	Measures	Explanation	Unit	Range
	Eye-gaze			
39	pupilMeanL	Average value of the left pupil diameter of the driver		
40	pupilMeanR	Average value of the right pupil diameter of the driver		
41	pupilStdL	Standard deviation of the left pupil diameter of the driver		
42	pupilStdR	Standard deviation of the right pupil dameter of the driver		
43	gazeSpeedMean	Average value of the gaze center moving speed of the driver		
44	gazeSpeedStd	Standard deviation of the gaze center moving speed of the driver		
45	gazeSpeedMax	Maximum value of the gaze center moving speed of the driver		
46	gazeDistanceCurrent	Gaze distance from a hazard at a pause		
47	gazeDistanceMin	Minimum gaze distance from a hazard within the time window		
48	tElapse	Elapsed time since the last fixation on a hazard at a pause		
49	tDwell	Fixation dwell time on a hazard within the time window		

Chapter 5

COMPARATIVE ASSESSMENT OF THE PERFORMANCE AND INTERPRETABILITY OF MODELS

5.1 Introduction

In the final phase of my research, I evaluated the model performance and the interpretability of the developed driver SA predictive models. While numerous studies have focused on the development of such models, a limited number have undertaken the crucial task of comparing the effectiveness and interpretability of multiple models. In this work, I propose the creation of robust SA predictive models while underscoring the evaluation and comparison of their interpretability.

The evaluation of model performance was based on general metrics widely recognized in the field. Key performance metrics include Accuracy, Precision, Recall, and F-1 score. These metrics provide a comprehensive assessment of the predictive power and effectiveness of each model in capturing different aspects of SA.

In addition to performance metrics, the interpretability of each model was a focal point of evaluation. SHapley Additive exPlanations (SHAP) was applied to assess the interpretability of the models. SHAP provides local explanations for individual predictions, offering insights into model decision-making at the instance level, as well as a global perspective, attributing the contribution of each feature to the model's output.

It is crucial to acknowledge that models inherently differ in their interpretability due to their distinct architectures and mechanisms. This inherent diversity was a critical aspect considered in my evaluation, ensuring a nuanced understanding of the interpretability spectrum across the chosen models.

In this chapter, I propose a machine learning model to predict driver SA at intersections by utilizing pedestrian interaction features, driving performance measures, and driver behavior data. Our approach employs eXtreme Gradient Boosting (XGBoost) combined with a novel feature selection method that integrates SHAP (SHapley Additive exPlanations) with the SelectKBest technique based on ANOVA F-values.

The contributions of this chapter are outlined as follows:

- We developed a real-time prediction model for the three levels of SA in conditionally automated driving.

- We introduced a feature selection framework that combines the SHAP model explanation technique with the SelectKBest ANOVA F-value approach. This framework consistently outperforms standalone methods in model accuracy across different feature set sizes.

- The model identified the most significant features for predicting SA. This includes the distance of pedestrians and the ego car from the intersection center, and the average and maximum speed of the ego car.

5.2 SHAP

SHapley Additive exPlanations (SHAP) encompasses main packages, namely KernelExplainer, TreeExplainer, and DeepExplainer. Notably, KernelExplainer and DeepExplainer may exhibit slowness when explaining all instances in large data sets, whereas TreeExplainer demonstrates efficiency in computing Shapley values for entire data sets, enabling robust global feature importance calculation.

The decomposed attributions provided by SHAP exhibit enhanced accuracy compared to those offered by Local Interpretable Model-Agnostic Explanations (LIME). Especially in simpler tasks, SHAP aligns more closely with human intuition, contributing to a more intuitive and understandable interpretability framework. SHAP boasts commendable mathematical properties, including consistency, missingness, and local accuracy, as highlighted in [78].

The SHAP value of a feature serves as a valuable metric to showcase the importance of each feature in the predictive model. This metric aids in understanding the influence of individual features on the overall prediction.

In the study conducted by [144], SHAP values of individual predictor variables in the LightGBM model were utilized to comprehend the factors influencing Situation Awareness (SA) and their respective effects. These SHAP values provided insights into the prediction model by identifying the most crucial factors and how they impact SA.

In the context of each model in our study, SHAP was employed for three primary purposes: 1) to identify the importance of individual predictor variables globally, including its use for feature selection, 2) to explain the main effects of important predictor variables on SA, and 3) to elucidate individual prediction instances by identifying the contributions of individual predictor variable-value sets [144].

5.3 Method

5.3.1 Predictive Model

To predict SA as a multi-class classification problem, we employed XGBoost, a gradient boosting framework that incorporates regularization within its loss function. The objective function utilized in XGBoost consists of two components: the cross-entropy loss function $L(\theta)$ and a regularization term $\Omega(\theta)$, which can be either $L1$ or $L2$ regularization. The regularization term $\Omega(\theta)$ serves to penalize model complexity, thus enhancing generalization performance.

Other models considered include logit regression, SVM, decision tree, random forest and adaboost. Table 5.1, 5.2, 5.3 and Figure 5.1 shows the comparative analysis of performance across the models. While XGBoost has a slightly larger spread (interquartile range) compared to Random Forest, its upper range extends to higher performance levels. In addition, XGBoost has a more consistent performance with fewer outliers.

Table 5.1: Performance metrics of various classifiers across 10 folds with 10 features

Fold	Method	Accuracy	F1	Precision	Recall
1	logit	0.5628	0.5600	0.5654	0.5628
1	svm	0.5628	0.5593	0.5652	0.5612
1	decision_tree	0.5410	0.5356	0.5356	0.5360
1	random_forest	0.6066	0.6021	0.6063	0.6022
1	xgboost	0.6120	0.6072	0.6120	0.6054
1	adaboost	0.5847	0.5726	0.5824	0.5716
2	logit	0.6011	0.5938	0.5942	0.5950
2	svm	0.6284	0.6219	0.6248	0.6210
2	decision_tree	0.5410	0.5228	0.5227	0.5235
2	random_forest	0.5628	0.5507	0.5521	0.5507
2	xgboost	0.5301	0.5171	0.5171	0.5172
2	adaboost	0.6557	0.6421	0.6487	0.6408
3	logit	0.5574	0.5477	0.5514	0.5467
3	svm	0.5792	0.5740	0.5797	0.5748
3	decision_tree	0.5082	0.4883	0.4875	0.4893
3	random_forest	0.5574	0.5455	0.5458	0.5458
3	xgboost	0.5792	0.5638	0.5644	0.5641
3	adaboost	0.5902	0.5770	0.5787	0.5773
4	logit	0.5847	0.5832	0.5868	0.5877
4	svm	0.6066	0.6061	0.6084	0.6113
4	decision_tree	0.5792	0.5641	0.5663	0.5641
4	random_forest	0.6612	0.6569	0.6587	0.6569
4	xgboost	0.5738	0.5676	0.5699	0.5689
4	adaboost	0.5847	0.5772	0.5774	0.5775
5	logit	0.5956	0.5860	0.5873	0.5858
5	svm	0.6011	0.5914	0.5926	0.5912
5	decision_tree	0.6120	0.6043	0.6055	0.6048
5	random_forest	0.6393	0.6295	0.6299	0.6302

Continued on next page

Table 5.1: (continued)

Fold	Method	Accuracy	F1	Precision	Recall
5	xgboost	0.6503	0.6415	0.6424	0.6418
5	adaboost	0.6284	0.6140	0.6146	0.6149
6	logit	0.5355	0.5330	0.5440	0.5345
6	svm	0.5628	0.5612	0.5734	0.5618
6	decision_tree	0.5027	0.4931	0.4969	0.4920
6	random_forest	0.5738	0.5690	0.5738	0.5692
6	xgboost	0.5301	0.5295	0.5324	0.5325
6	adaboost	0.5683	0.5622	0.5645	0.5624
7	logit	0.5628	0.5620	0.5720	0.5658
7	svm	0.5683	0.5679	0.5822	0.5715
7	decision_tree	0.4863	0.4735	0.4744	0.4742
7	random_forest	0.5464	0.5436	0.5462	0.5451
7	xgboost	0.5738	0.5667	0.5684	0.5665
7	adaboost	0.5683	0.5619	0.5625	0.5619
8	logit	0.5464	0.5440	0.5491	0.5496
8	svm	0.5902	0.5883	0.5914	0.5985
8	decision_tree	0.5464	0.5432	0.5485	0.5448
8	random_forest	0.5956	0.5906	0.5903	0.5912
8	xgboost	0.5738	0.5701	0.5716	0.5727
8	adaboost	0.5738	0.5713	0.5717	0.5734
9	logit	0.5628	0.5502	0.5557	0.5513
9	svm	0.5464	0.5383	0.5440	0.5364
9	decision_tree	0.5410	0.5346	0.5388	0.5348
9	random_forest	0.5683	0.5630	0.5690	0.5634
9	xgboost	0.5574	0.5458	0.5469	0.5460
9	adaboost	0.5628	0.5501	0.5543	0.5502
10	logit	0.5628	0.5607	0.5598	0.5645
10	svm	0.5792	0.5716	0.5763	0.5794

Continued on next page

Table 5.1: (continued)

Fold	Method	Accuracy	F1	Precision	Recall
10	decision_tree	0.5574	0.5518	0.5518	0.5522
10	random_forest	0.6011	0.5850	0.5883	0.5858
10	xgboost	0.5902	0.5800	0.5810	0.5794
10	adaboost	0.5519	0.5318	0.5368	0.5346

Table 5.2: Performance metrics of various classifiers across 10 folds with 20 features

Fold	Method	Accuracy	F1	Precision	Recall
1	logit	0.5956	0.5912	0.5926	0.5927
1	svm	0.5847	0.5783	0.5805	0.5787
1	decision_tree	0.5246	0.5170	0.5177	0.5172
1	random_forest	0.6011	0.5947	0.5984	0.5927
1	xgboost	0.6011	0.5930	0.5959	0.5920
1	adaboost	0.5574	0.5452	0.5494	0.5453
2	logit	0.6066	0.5950	0.5951	0.5950
2	svm	0.6175	0.6044	0.6109	0.6036
2	decision_tree	0.5410	0.5296	0.5302	0.5305
2	random_forest	0.5956	0.5755	0.5770	0.5770
2	xgboost	0.5355	0.5200	0.5197	0.5208
2	adaboost	0.6503	0.6352	0.6396	0.6344
3	logit	0.5628	0.5492	0.5499	0.5491
3	svm	0.5738	0.5644	0.5648	0.5643
3	decision_tree	0.4918	0.4667	0.4649	0.4700
3	random_forest	0.5628	0.5416	0.5446	0.5433
3	xgboost	0.5792	0.5567	0.5614	0.5585
3	adaboost	0.5902	0.5714	0.5776	0.5732
4	logit	0.5956	0.5940	0.5952	0.5981

Continued on next page

Table 5.2: (continued)

Fold	Method	Accuracy	F1	Precision	Recall
4	svm	0.5956	0.5957	0.5974	0.5993
4	decision_tree	0.5301	0.5276	0.5296	0.5313
4	random_forest	0.6011	0.5976	0.5988	0.5979
4	xgboost	0.6175	0.6122	0.6158	0.6118
4	adaboost	0.5956	0.5891	0.5893	0.5904
5	logit	0.6230	0.6051	0.6100	0.6055
5	svm	0.6284	0.6163	0.6197	0.6154
5	decision_tree	0.6284	0.6227	0.6241	0.6230
5	random_forest	0.6448	0.6358	0.6382	0.6347
5	xgboost	0.6557	0.6503	0.6546	0.6480
5	adaboost	0.6393	0.6266	0.6275	0.6267
6	logit	0.5683	0.5596	0.5617	0.5591
6	svm	0.5519	0.5472	0.5547	0.5487
6	decision_tree	0.5792	0.5792	0.5811	0.5856
6	random_forest	0.6011	0.5942	0.5973	0.5940
6	xgboost	0.5792	0.5728	0.5797	0.5721
6	adaboost	0.5847	0.5815	0.5853	0.5817
7	logit	0.5738	0.5696	0.5722	0.5710
7	svm	0.5410	0.5386	0.5452	0.5389
7	decision_tree	0.5082	0.5024	0.5030	0.5034
7	random_forest	0.5847	0.5796	0.5840	0.5801
7	xgboost	0.5956	0.5896	0.5948	0.5903
7	adaboost	0.5410	0.5371	0.5419	0.5393
8	logit	0.5246	0.5186	0.5257	0.5273
8	svm	0.5792	0.5768	0.5769	0.5812
8	decision_tree	0.5355	0.5171	0.5226	0.5205
8	random_forest	0.6120	0.6014	0.6057	0.6029
8	xgboost	0.5628	0.5534	0.5544	0.5568

Continued on next page

Table 5.2: (continued)

Fold	Method	Accuracy	F1	Precision	Recall
8	adaboost	0.5464	0.5412	0.5476	0.5445
9	logit	0.5464	0.5313	0.5358	0.5327
9	svm	0.5792	0.5656	0.5658	0.5661
9	decision_tree	0.5137	0.5063	0.5079	0.5074
9	random_forest	0.6066	0.6006	0.6023	0.6013
9	xgboost	0.6503	0.6378	0.6381	0.6378
9	adaboost	0.5902	0.5761	0.5757	0.5769
10	logit	0.5792	0.5716	0.5781	0.5731
10	svm	0.5574	0.5503	0.5550	0.5534
10	decision_tree	0.5847	0.5783	0.5794	0.5783
10	random_forest	0.6557	0.6411	0.6443	0.6415
10	xgboost	0.6448	0.6308	0.6310	0.6311
10	adaboost	0.6230	0.6161	0.6164	0.6165

Table 5.3: Performance metrics of various classifiers across 10 folds with 30 features

Fold	Method	Accuracy	F1	Precision	Recall
1	logit	0.6120	0.6065	0.6062	0.6074
1	svm	0.6393	0.6273	0.6300	0.6265
1	decision_tree	0.5410	0.5383	0.5392	0.5406
1	random_forest	0.6066	0.6013	0.6056	0.5995
1	xgboost	0.6393	0.6285	0.6362	0.6267
1	adaboost	0.5956	0.5842	0.5861	0.5852
2	logit	0.5847	0.5763	0.5786	0.5764
2	svm	0.6284	0.6187	0.6231	0.6169
2	decision_tree	0.5410	0.5248	0.5307	0.5278
2	random_forest	0.6066	0.5920	0.5934	0.5922

Continued on next page

Table 5.3: (continued)

Fold	Method	Accuracy	F1	Precision	Recall
2	xgboost	0.5628	0.5487	0.5499	0.5491
2	adaboost	0.6175	0.6036	0.6051	0.6033
3	logit	0.5628	0.5471	0.5496	0.5484
3	svm	0.5683	0.5513	0.5534	0.5523
3	decision_tree	0.5137	0.5023	0.5032	0.5025
3	random_forest	0.5628	0.5429	0.5439	0.5442
3	xgboost	0.6120	0.5885	0.5928	0.5904
3	adaboost	0.5847	0.5706	0.5721	0.5716
4	logit	0.6011	0.5990	0.6005	0.6029
4	svm	0.6120	0.6111	0.6119	0.6120
4	decision_tree	0.5301	0.5249	0.5296	0.5279
4	random_forest	0.6230	0.6206	0.6229	0.6231
4	xgboost	0.6120	0.6064	0.6100	0.6065
4	adaboost	0.6066	0.6043	0.6094	0.6088
5	logit	0.6284	0.6127	0.6188	0.6129
5	svm	0.6612	0.6439	0.6501	0.6436
5	decision_tree	0.5902	0.5840	0.5835	0.5864
5	random_forest	0.6503	0.6426	0.6439	0.6418
5	xgboost	0.6612	0.6538	0.6545	0.6537
5	adaboost	0.6120	0.5977	0.5980	0.5978
6	logit	0.5792	0.5717	0.5722	0.5713
6	svm	0.6066	0.5968	0.6001	0.5957
6	decision_tree	0.5574	0.5546	0.5629	0.5550
6	random_forest	0.6284	0.6231	0.6254	0.6227
6	xgboost	0.6175	0.6117	0.6167	0.6107
6	adaboost	0.6066	0.5910	0.5947	0.5907
7	logit	0.5519	0.5435	0.5446	0.5432
7	svm	0.6230	0.6129	0.6157	0.6143

Continued on next page

Table 5.3: (continued)

Fold	Method	Accuracy	F1	Precision	Recall
7	decision_tree	0.5082	0.5065	0.5054	0.5096
7	random_forest	0.5792	0.5755	0.5793	0.5769
7	xgboost	0.5902	0.5842	0.5843	0.5852
7	adaboost	0.5628	0.5591	0.5593	0.5601
8	logit	0.5628	0.5540	0.5643	0.5590
8	svm	0.5847	0.5779	0.5850	0.5842
8	decision_tree	0.5082	0.4909	0.4918	0.4934
8	random_forest	0.6120	0.6037	0.6070	0.6039
8	xgboost	0.5902	0.5776	0.5811	0.5795
8	adaboost	0.5628	0.5505	0.5523	0.5524
9	logit	0.5464	0.5309	0.5311	0.5324
9	svm	0.5956	0.5869	0.5886	0.5865
9	decision_tree	0.5410	0.5352	0.5420	0.5352
9	random_forest	0.6230	0.6158	0.6195	0.6154
9	xgboost	0.6120	0.5988	0.6017	0.5984
9	adaboost	0.6066	0.5806	0.5798	0.5848
10	logit	0.5628	0.5519	0.5557	0.5550
10	svm	0.5683	0.5564	0.5673	0.5605
10	decision_tree	0.6120	0.6027	0.6037	0.6036
10	random_forest	0.6393	0.6289	0.6336	0.6285
10	xgboost	0.6393	0.6265	0.6283	0.6266
10	adaboost	0.5792	0.5665	0.5752	0.5669

5.3.2 Feature Selection

Features were organized into various groups, including overlap features, along with additional feature sets defined in the features module, such as shap, kbest, and combination. This systematic selection of features aimed to leverage diverse insights for the model. To identify the most predictive features contributing to driver situational awareness (SA), we conducted a comparative analysis using two distinct feature selection

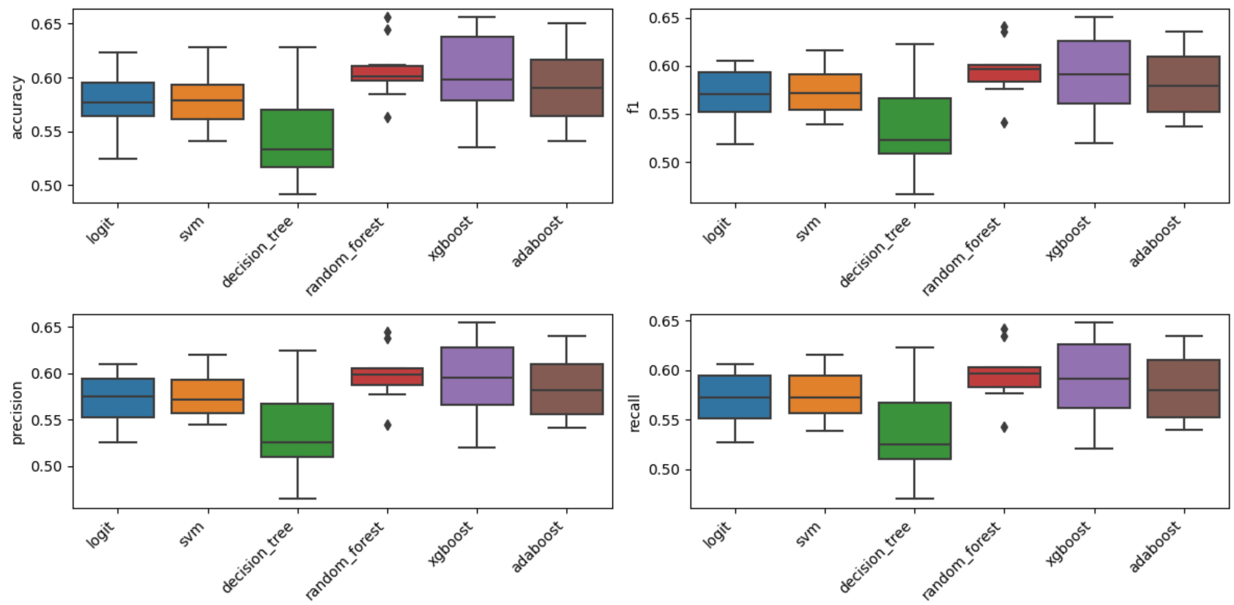


Figure 5.1: Base performance comparative analysis of different predictive models of driver SA

techniques: ANOVA and SHAP. The results of these methods, as well as a combined approach, are presented in Figure X.

The ANOVA method selects features based on their ability to explain between-group variance in SA labels. While it provides a straightforward statistical basis for feature relevance, it does not account for potential nonlinear interactions or contextual dependencies within the data. Conversely, the SHAP method evaluates feature importance by quantifying the marginal contribution of each feature to the model's prediction, allowing for both nonlinear and model-specific insights.

The top 10 features from each method are shown in order of importance. Notably, several features—such as CenterDistanceTotal, ObjectCorner, and CenterDistanceY—appear in multiple methods, suggesting they are robust predictors of SA. Features common across methods are bolded, while the color coding indicates the source method: red for ANOVA, blue for SHAP, and purple for overlap.

The combined feature set aggregates the most informative features from both methods, aiming to retain predictive power while enhancing generalizability. This hybrid selection is particularly valuable when deploying models in real-world scenarios where both linear and nonlinear patterns may influence driver awareness.

These results highlight the benefit of integrating interpretable feature selection techniques to ensure both statistical rigor and practical relevance in modeling human situational awareness.

5.3.3 Model Evaluation

The model's performance was assessed using various metrics, including accuracy, F1 score, precision, and recall, calculated across the outer validation folds. These metrics provide a holistic view of the model's predictive capability and its balance between precision and recall.

5.4 Model Performance

Table 5.4, 5.5, 5.6 and Figure 5.2 shows the comparative analysis between three feature selection methods – SHAP, ANOVA, and a combined method – across four classification metrics: Accuracy, Macro F1 score, Precision, and Recall. Each sub-plot shows the performance distributions of the methods with varying numbers of selected features (5, 10, 15, 20, 25, and 30).

Table 5.4: Model Performance Metrics by Number of Features using ANOVA

Accuracy	F1	Precision	Recall	Method	Num Features
0.6503	0.6382	0.6386	0.6387	anova	5
0.6011	0.5825	0.5822	0.5829	anova	5
0.5792	0.5698	0.5693	0.5713	anova	5
0.6011	0.5899	0.5901	0.5898	anova	5
0.5792	0.5684	0.5697	0.5684	anova	5
0.5847	0.5747	0.5784	0.5739	anova	5
0.5683	0.5584	0.5584	0.5586	anova	5
0.5464	0.5376	0.5421	0.5367	anova	5
0.5355	0.5233	0.5258	0.5264	anova	5
0.541	0.5416	0.552	0.544	anova	5
0.612	0.605	0.6063	0.6043	anova	10
0.5847	0.5713	0.5743	0.5709	anova	10
0.5519	0.5388	0.5398	0.5387	anova	10
0.6011	0.5824	0.5841	0.5829	anova	10
0.5847	0.5808	0.5817	0.5801	anova	10

Accuracy	F1	Precision	Recall	Method	Num Features
0.5847	0.574	0.5744	0.5739	anova	10
0.5792	0.5728	0.5727	0.5737	anova	10
0.5355	0.529	0.5347	0.5289	anova	10
0.5956	0.5869	0.5905	0.5861	anova	10
0.612	0.6073	0.6062	0.6098	anova	10
0.5738	0.5619	0.5663	0.5621	anova	15
0.5738	0.5663	0.5686	0.5659	anova	15
0.5738	0.5705	0.5742	0.5702	anova	15
0.6503	0.637	0.6413	0.6362	anova	15
0.6339	0.6238	0.6281	0.6249	anova	15
0.5956	0.5839	0.5867	0.5835	anova	15
0.5683	0.5541	0.5548	0.554	anova	15
0.6011	0.5897	0.5927	0.5906	anova	15
0.6066	0.5844	0.5869	0.587	anova	15
0.5792	0.5722	0.5768	0.5742	anova	15
0.5792	0.566	0.5681	0.5655	anova	20
0.6448	0.6345	0.6352	0.6349	anova	20
0.5683	0.5634	0.5671	0.5623	anova	20
0.6175	0.6134	0.6136	0.6156	anova	20
0.6066	0.5892	0.5899	0.5898	anova	20
0.6066	0.5966	0.5968	0.5966	anova	20
0.623	0.6129	0.6228	0.6102	anova	20
0.5738	0.5656	0.5669	0.5657	anova	20
0.612	0.596	0.6082	0.5955	anova	20
0.5956	0.588	0.588	0.5883	anova	20
0.6339	0.6272	0.6336	0.6242	anova	25
0.5956	0.5863	0.5869	0.5863	anova	25
0.612	0.6011	0.6013	0.6011	anova	25
0.5956	0.5841	0.5848	0.5838	anova	25
0.5956	0.5763	0.5862	0.578	anova	25
0.6284	0.6188	0.6206	0.6187	anova	25

Accuracy	F1	Precision	Recall	Method	Num Features
0.694	0.6896	0.6893	0.69	anova	25
0.5902	0.5815	0.5834	0.5809	anova	25
0.612	0.5941	0.5938	0.5951	anova	25
0.612	0.6014	0.6046	0.6006	anova	25
0.5628	0.544	0.5479	0.5464	anova	30
0.5956	0.5855	0.5856	0.5854	anova	30
0.5956	0.593	0.5924	0.5956	anova	30
0.6393	0.6299	0.6296	0.6303	anova	30
0.6831	0.6796	0.6805	0.6797	anova	30
0.5902	0.5788	0.583	0.5773	anova	30
0.6011	0.5918	0.5923	0.5915	anova	30
0.5956	0.5782	0.5785	0.5803	anova	30
0.6284	0.62	0.6218	0.6199	anova	30
0.6011	0.5838	0.6015	0.5832	anova	30

Table 5.5: Model Performance Metrics by Number of Features using SHAP

Accuracy	F1	Precision	Recall	Method	Num Features
0.5956	0.5787	0.5824	0.5791	shap	5
0.5683	0.5531	0.5534	0.5535	shap	5
0.4973	0.4853	0.4854	0.4852	shap	5
0.5519	0.5439	0.5479	0.5424	shap	5
0.6011	0.5905	0.5922	0.5903	shap	5
0.4809	0.478	0.4798	0.4794	shap	5
0.5847	0.5665	0.5719	0.5669	shap	5
0.5628	0.5547	0.558	0.5538	shap	5
0.612	0.6038	0.6038	0.6039	shap	5
0.5628	0.5426	0.5424	0.5454	shap	5
0.5574	0.5464	0.5453	0.5487	shap	10
0.5519	0.5367	0.5377	0.5385	shap	10

Accuracy	F1	Precision	Recall	Method	Num Features
0.5902	0.576	0.5774	0.5754	shap	10
0.5082	0.5009	0.5018	0.5004	shap	10
0.5464	0.5301	0.53	0.5305	shap	10
0.5847	0.5724	0.575	0.5719	shap	10
0.6066	0.5884	0.5911	0.5899	shap	10
0.5792	0.5662	0.568	0.5664	shap	10
0.5902	0.573	0.5747	0.5743	shap	10
0.5956	0.5891	0.5912	0.5957	shap	10
0.6175	0.6087	0.6187	0.6054	shap	15
0.623	0.6143	0.6144	0.6147	shap	15
0.6066	0.5896	0.5959	0.589	shap	15
0.5027	0.4975	0.4986	0.4975	shap	15
0.6393	0.6364	0.6384	0.6385	shap	15
0.5738	0.5647	0.5657	0.5642	shap	15
0.6393	0.6241	0.637	0.6263	shap	15
0.5847	0.5636	0.5668	0.5654	shap	15
0.6175	0.6023	0.6056	0.6025	shap	15
0.6393	0.6323	0.6332	0.6332	shap	15
0.5956	0.5662	0.5897	0.5696	shap	20
0.5464	0.5412	0.5412	0.5417	shap	20
0.623	0.612	0.6142	0.611	shap	20
0.5847	0.5604	0.5627	0.5648	shap	20
0.6175	0.608	0.6079	0.6083	shap	20
0.6393	0.6287	0.6314	0.6276	shap	20
0.623	0.6029	0.6115	0.6041	shap	20
0.6011	0.5875	0.5925	0.5876	shap	20
0.6339	0.6269	0.6276	0.6273	shap	20
0.6066	0.597	0.603	0.5954	shap	20
0.6066	0.5908	0.5914	0.5924	shap	25
0.612	0.605	0.6079	0.6051	shap	25
0.623	0.6203	0.6227	0.6221	shap	25

Accuracy	F1	Precision	Recall	Method	Num Features
0.5956	0.5702	0.586	0.5729	shap	25
0.6284	0.6241	0.6268	0.6241	shap	25
0.6557	0.642	0.6478	0.641	shap	25
0.6448	0.6289	0.6418	0.6271	shap	25
0.6393	0.6342	0.6373	0.6328	shap	25
0.6011	0.5843	0.5932	0.584	shap	25
0.6011	0.5806	0.5991	0.5814	shap	25
0.6776	0.6671	0.6666	0.6679	shap	30
0.623	0.6011	0.6007	0.6031	shap	30
0.5902	0.5897	0.5883	0.5925	shap	30
0.6557	0.6426	0.6545	0.6401	shap	30
0.6011	0.5904	0.5942	0.5911	shap	30
0.5956	0.586	0.5878	0.5855	shap	30
0.6175	0.5884	0.601	0.5925	shap	30
0.6393	0.62	0.6298	0.6223	shap	30
0.6284	0.6088	0.6153	0.61	shap	30
0.5683	0.5624	0.5646	0.5612	shap	30

Table 5.6: Model Performance Metrics by Number of Features using Combined Method

Accuracy	F1	Precision	Recall	Method	Num Features
0.623	0.6173	0.6175	0.6183	combined	10
0.5738	0.5582	0.5596	0.5582	combined	10
0.5738	0.5625	0.5629	0.5628	combined	10
0.5956	0.5931	0.5974	0.5918	combined	10
0.6612	0.6529	0.6554	0.6519	combined	10
0.5137	0.5103	0.5195	0.514	combined	10
0.612	0.6012	0.6014	0.6012	combined	10
0.5847	0.582	0.5836	0.5841	combined	10
0.5956	0.5779	0.5814	0.5791	combined	10

Accuracy	F1	Precision	Recall	Method	Num Features
0.6066	0.5889	0.5949	0.5896	combined	10
0.6448	0.6374	0.6469	0.6337	combined	20
0.6175	0.6093	0.6093	0.6099	combined	20
0.6503	0.6389	0.6409	0.6385	combined	20
0.5191	0.509	0.5102	0.5091	combined	20
0.5792	0.5587	0.5658	0.5598	combined	20
0.612	0.5972	0.5998	0.5969	combined	20
0.623	0.5988	0.6006	0.6017	combined	20
0.5847	0.5653	0.5693	0.5672	combined	20
0.6448	0.6393	0.6428	0.6381	combined	20
0.6284	0.6225	0.6237	0.6221	combined	20
0.6175	0.6035	0.6098	0.6022	combined	30
0.6557	0.6362	0.643	0.6364	combined	30
0.5847	0.5682	0.5706	0.5691	combined	30
0.6667	0.663	0.6628	0.6634	combined	30
0.694	0.6844	0.6883	0.6827	combined	30
0.6339	0.6255	0.6276	0.6244	combined	30
0.612	0.5999	0.6006	0.5995	combined	30
0.612	0.5903	0.5974	0.5944	combined	30
0.6448	0.6265	0.6298	0.6289	combined	30
0.623	0.6135	0.6135	0.6136	combined	30

Overall, the combined feature selection method exhibits a balanced performance across all four metrics, especially as the number of features increases. This method appears to mitigate the variability observed with SHAP and ANOVA, potentially offering a more robust solution for optimal feature selection in this classification task. SHAP, however, demonstrates superior performance in precision and recall with lower feature counts, suggesting it may be more suitable when aiming to reduce dimensionality without sacrificing specific metrics. Conversely, ANOVA consistently underperforms compared to SHAP and the combined method, making it less favorable for tasks that require high accuracy and consistency across various feature counts.

5.4.1 Accuracy

The SHAP method generally demonstrates a broader range of accuracy, with relatively higher median values as the number of features increases. The combined method appears to achieve more consistent accuracy, particularly for higher feature counts (25 and 30), where it achieves the highest median accuracy. ANOVA shows a narrower range of accuracy distributions compared to SHAP, but it also tends to under perform, especially at lower feature counts (5 and 10).

5.4.2 F1 Score

Across feature counts, SHAP and the combined method perform comparably in terms of F1 score, with SHAP showing slightly higher median values in some cases. Similar to accuracy, ANOVA generally yields lower F1 scores, particularly when fewer features are selected. The combined method shows a tighter distribution as the feature count increases, indicating potential robustness at higher feature counts.

5.4.3 Precision

SHAP consistently demonstrates a higher range of precision values, especially with fewer features (5, 10, and 15), suggesting that SHAP may better capture the relevant features for precision at lower feature counts. ANOVA lags in performance for precision, particularly in lower feature counts, and shows more variability across counts. The combined method yields a balanced precision distribution as the number of features increases, reaching comparable precision to SHAP at higher feature counts (25 and 30).

5.4.4 Recall

Recall scores are generally consistent across methods, with SHAP and the combined method showing a slight advantage, especially with a higher feature count. ANOVA demonstrates lower recall in some cases, particularly for lower feature counts (5 and 10), though it performs more consistently at higher counts. The combined method once again shows a balanced recall distribution across feature counts, with its performance aligning closely with SHAP at higher feature counts.

5.5 Interpretation

5.5.1 Global Model Interpretation

Figure 5.3 presents the global SHAP summary plot for the model employing the combined feature selection method, offering a comprehensive view of how each feature influences model predictions across the entire dataset. Each point on the plot corresponds to an individual instance, with color denoting the feature value (e.g., low to high), and position on the x-axis indicating the SHAP value associated with that instance. The SHAP value reflects the marginal contribution of a feature to the prediction: positive values indicate an increase in the predicted SA level, while negative values indicate a decrease.

The y-axis ranks the top ten most influential features, as determined by their mean absolute SHAP values. Notably, the distance from the pedestrian to the intersection center emerges as a key predictor. When this distance is low, the associated SHAP values tend to be strongly negative, indicating that close pedestrian proximity reduces the predicted SA level. In contrast, larger pedestrian distances are associated with highly positive SHAP values, suggesting a positive contribution to situational awareness. This finding aligns with theoretical expectations: closer pedestrian proximity may limit a driver's time to perceive and interpret the environment, thereby reducing SA, whereas more distant pedestrians afford the driver greater opportunity to process environmental cues, facilitating higher SA levels.

Overall, the SHAP summary plot provides both global feature importance and nuanced insights into the directionality and magnitude of feature effects, underscoring the interpretability of the predictive model developed using the combined feature selection framework.

One of the most influential features in the model is Pedestrian Distance from Intersection Center. The model generally predicts higher Situational Awareness (SA) scores when pedestrians are located farther from the intersection center (e.g., on the sidewalk), likely reflecting increased time and space for driver assessment. However, when pedestrians are near the center, the predicted SA scores exhibit greater variability — suggesting no clear linear relationship. This observation emphasizes the necessity for context-aware interpretation in close-proximity scenarios and highlights the limitations of traditional linear analyses, such as ANOVA, in capturing such nuanced patterns.

From an applied perspective, this feature plays a key role in determining when a driver alert should be triggered. By modeling these nonlinear effects, the system can support smarter and more appropriate alert timing.

The second most important feature, Ego Car Distance, demonstrates an effect on SA opposite to that of pedestrian distance. This inverse relationship suggests a complex interaction between the relative positions of pedestrians and the vehicle, reinforcing the need for models capable of capturing nonlinear and interactive dynamics.

5.5.2 Local Interpretation Examples of varying SA

Figure 5.4 illustrates the SHAP force plot for an individual instance, providing a localized explanation of the model's prediction for the Situational Awareness (SA) score. The features are ordered from top to bottom based on their absolute contribution to the prediction, with those exerting the greatest influence appearing at the top. Each feature is accompanied by its actual value for this specific data point, displayed in gray text adjacent to the corresponding bar.

The model prediction begins with a base value, which represents the mean predicted SA score across the entire dataset. From this baseline, the final predicted SA score for the instance is obtained by summing the SHAP values of all features. Each SHAP value quantifies the extent to which a feature contributes to increasing or decreasing the prediction.

In the plot, red bars indicate features that contribute positively—i.e., they push the prediction toward a higher SA score. Conversely, blue bars denote negative contributions, decreasing the predicted SA score. The length of each bar reflects the magnitude of the feature's influence, while its direction (left or right) indicates whether the effect is negative or positive, respectively.

This visualization offers a transparent, instance-specific decomposition of the model's decision-making process, thereby enhancing interpretability and supporting model trustworthiness in high-stakes decision contexts.

Figure 5.5 presents the SHAP force plot for a different instance characterized by a low predicted driver Situational Awareness (SA) score. Compared to the previous instance, the ranking of features differs, indicating that the relative importance of individual features has changed for this prediction. Moreover, the SHAP values associated with the same features vary in magnitude and direction, reflecting context-specific contributions. These differences ultimately result in a distinct predicted SA score, underscoring the model's sensitivity to feature interactions and instance-level variability.

This plot provides a local interpretation — it shows how each feature contributed to the predicted SA score for one specific instance in the dataset. Features are listed from top to bottom, with the most impactful ones at the top. Next to each feature, there's a gray number — which represents the actual value of that feature for this particular instance. The prediction starts from what's called a base value , which is the average prediction across the entire dataset. From there, each feature either pushes the prediction upward or downward, depending on its SHAP value. Red bars indicate features that increased the prediction — meaning they pushed the SA score higher. Blue bars indicate features that decreased the prediction — pulling the score lower. The length of each bar shows how strong that influence was. And the direction — whether the bar goes left or right — tells whether it was a positive or negative contribution. So this visualization show us why the model made the prediction it did, in a way that's intuitive and interpretable. You can see that the features are ranked differently in this prediction, which means that the impact of each feature has changed. Also, the same feature has different shap values compared to the previous prediction resulting in different final prediction.

5.6 Discussion

The differences in performance among the feature selection methods can be attributed to how each method identifies and ranks features, as well as the underlying assumptions each method makes about feature importance and relationships with the target variable.

SHAP values estimate the contribution of each feature to a model's predictions by considering all possible combinations of features. This method captures complex, non-linear relationships between features and the target variable, which can be advantageous in models that involve interactions and dependencies among features. The higher accuracy, precision, and recall observed with SHAP, especially for fewer features, suggest that SHAP identifies features that are more influential for these metrics in a nuanced way. However, this approach can also lead to more variability due to the sensitivity to outliers or overfitting when fewer features are selected.

The relatively lower performance of ANOVA, particularly for accuracy and F1 score, may result from its inability to capture interactions or complex dependencies. ANOVA may perform well in cases where the features are truly independent of each other and have direct linear relationships with the target, but this assumption often doesn't hold in real-world data, limiting its effectiveness in identifying the most predictive features.

The combined method likely leverages complementary strengths of both SHAP and ANOVA, leading to a more balanced performance across metrics [16, 126]. By combining the comprehensive, interaction-aware feature selection of SHAP with the simpler, interpretable ranking from ANOVA, this method appears to mitigate the limitations of each individual approach. As the number of features increases, the combined method seems to stabilize, potentially because it includes diverse types of important features (both linear and non-linear predictors) without introducing as much noise as either method alone might at higher feature counts.

SHAP analysis suggests that when the distance from the center of a pedestrian is low, the SHAP value is highly negative, and when the distance is high, the SHAP value is highly positive. A highly negative SHAP value means that a low pedestrian distance decreases the model's predicted SA score. This suggests that when a pedestrian is closer to the driver's central field of vision, the model predicts lower situational awareness. A pedestrian in close proximity may not provide enough time for drivers to comprehend and project (SA level 2 and 3). On the other hand, A highly positive SHAP value for greater pedestrian distance means that the model predicts higher SA when pedestrians are farther from the center. It may be due to when pedestrians are further away, this provides drivers enough time to comprehend the environment, allowing the driver to maintain better overall awareness of the driving environment.

This study investigated a range of environmental and behavioral factors influencing driver Situational

Awareness (SA) in complex urban scenarios. In particular, the analysis identified key pedestrian-related features and driving performance indicators that significantly affect SA predictions. A novel feature selection approach was proposed, incorporating model interpretability techniques (e.g., SHAP) to enhance the transparency and robustness of the selection process.

Comprehensive predictive models of driver SA were developed and systematically interpreted, enabling deeper insights into the relative impact and behavioral implications of individual features. The integration of interpretability-driven insights further led to improved model performance, demonstrating the practical utility of explainable AI approaches in human factors modeling.

Despite these advancements, several limitations remain. First, the current model requires integration with additional in-vehicle sensors to enable robust real-time SA prediction under varying driving conditions. Second, the computational complexity of the interpretability methods may present challenges for real-time, on-board implementation in resource-constrained vehicle systems. Finally, while the model shows strong predictive capabilities, further validation using diverse and real-world driving data is necessary to ensure generalizability and deployment readiness.

Looking ahead, there are several directions for future research that could advance this work./ First, the relationship between model interpretability and human trust needs further investigation. A key question is how the transparency of predictive models influences driver trust in autonomous systems. For example, how much information about a vehicle's decision-making process is optimal for building trust? Additionally, it would be insightful to examine whether trust levels differ between novice and experienced drivers when interacting with autonomous vehicle features.

Another valuable work involves validating the combined feature selection method of ANOVA and SHAP. While this method demonstrated effectiveness in my thesis, testing their capabilities across diverse datasets could provide valuable insights into their generalizability and robustness.

Lastly, a critical next step would be to validate these findings using real-world driving data. The data for this dissertation was collected in a driving simulator, which allowed for controlled testing conditions. However, applying these models to actual driving scenarios would provide a more comprehensive assessment of their performance under real-world conditions.

5.7 Conclusion

In conclusion, this chapter highlights the comparative analysis of multiple predictive models and feature selection techniques for modeling driver SA. Among the evaluated models, XGBoost was found to be the most robust in terms of performance and consistency. Furthermore, the proposed hybrid feature selection method, combining SHAP and ANOVA, demonstrated superior and more stable performance across various evaluation metrics compared to either method used independently. SHAP, in particular, provided valuable interpretability at both global and local levels, uncovering key insights into how feature values, such as pedestrian distance from the intersection center, influence SA predictions. These findings not only enhance the predictive accuracy of SA models but also offer interpretable insights critical for deploying intelligent driver assistance systems in real-world contexts.

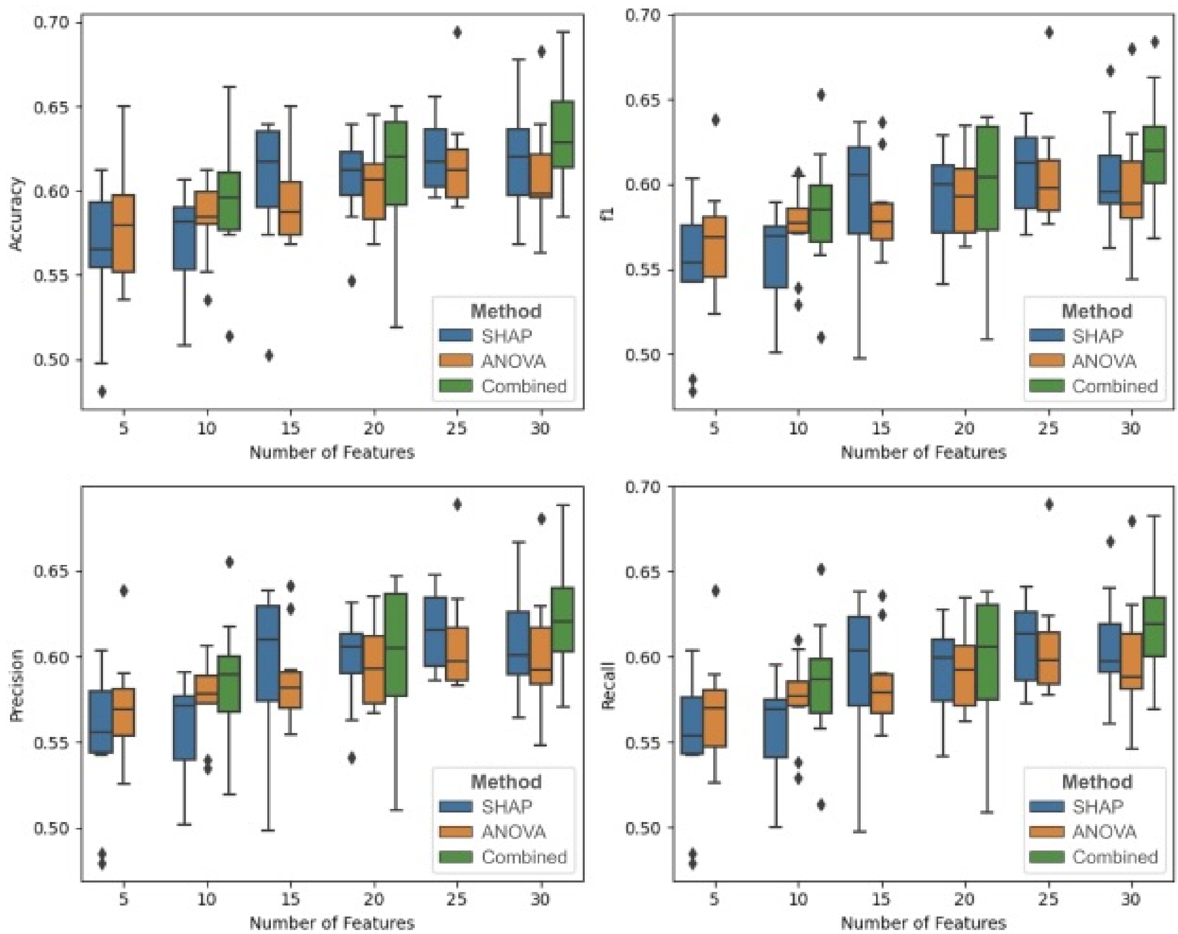


Figure 5.2: Comparative analysis between three feature selection methods

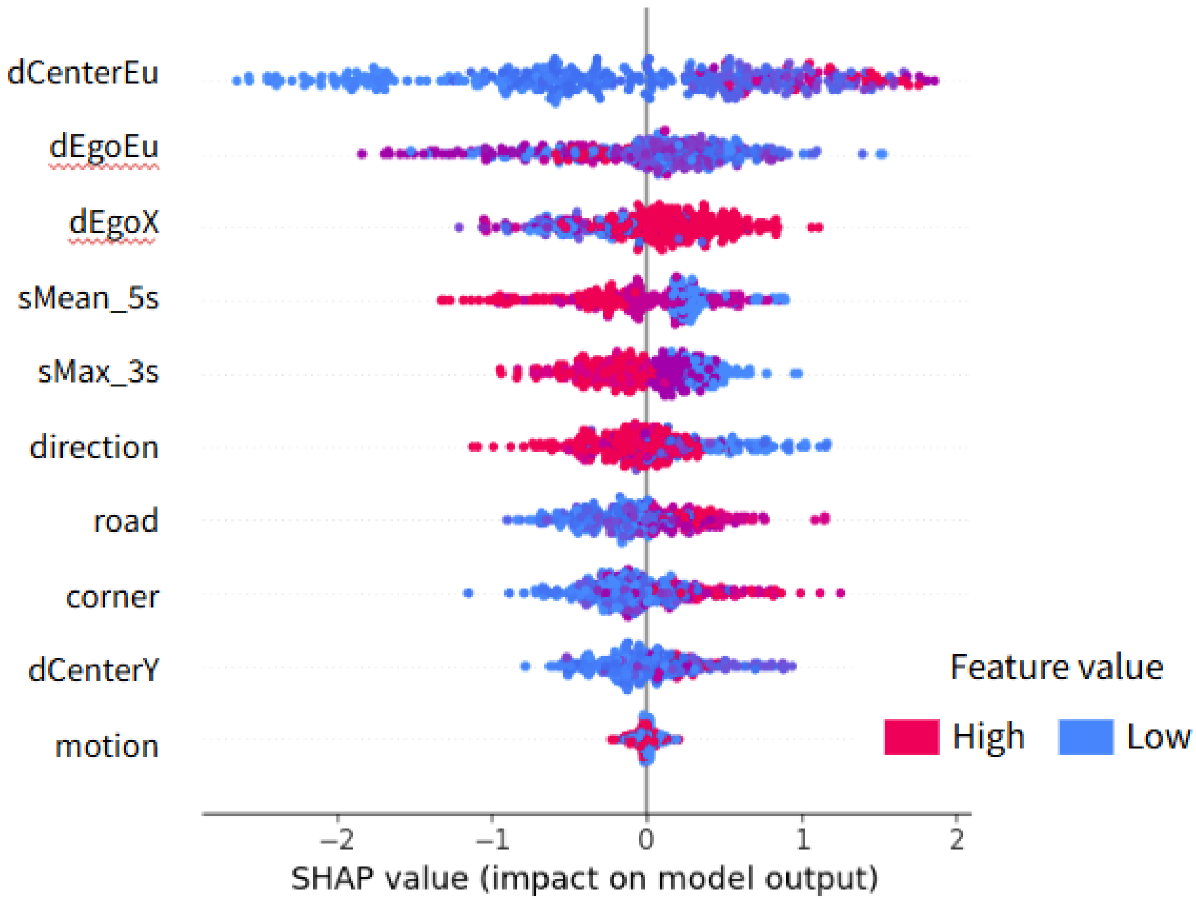


Figure 5.3: Global SHAP summary plot

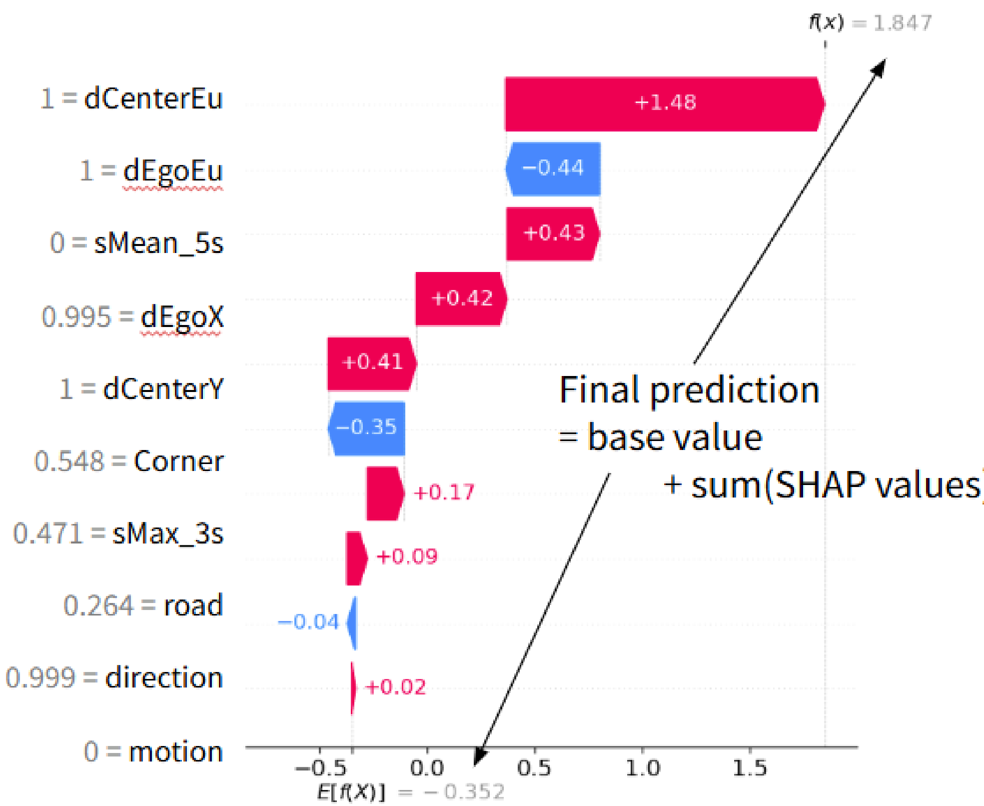


Figure 5.4: Local SHAP summary plot of high driver SA

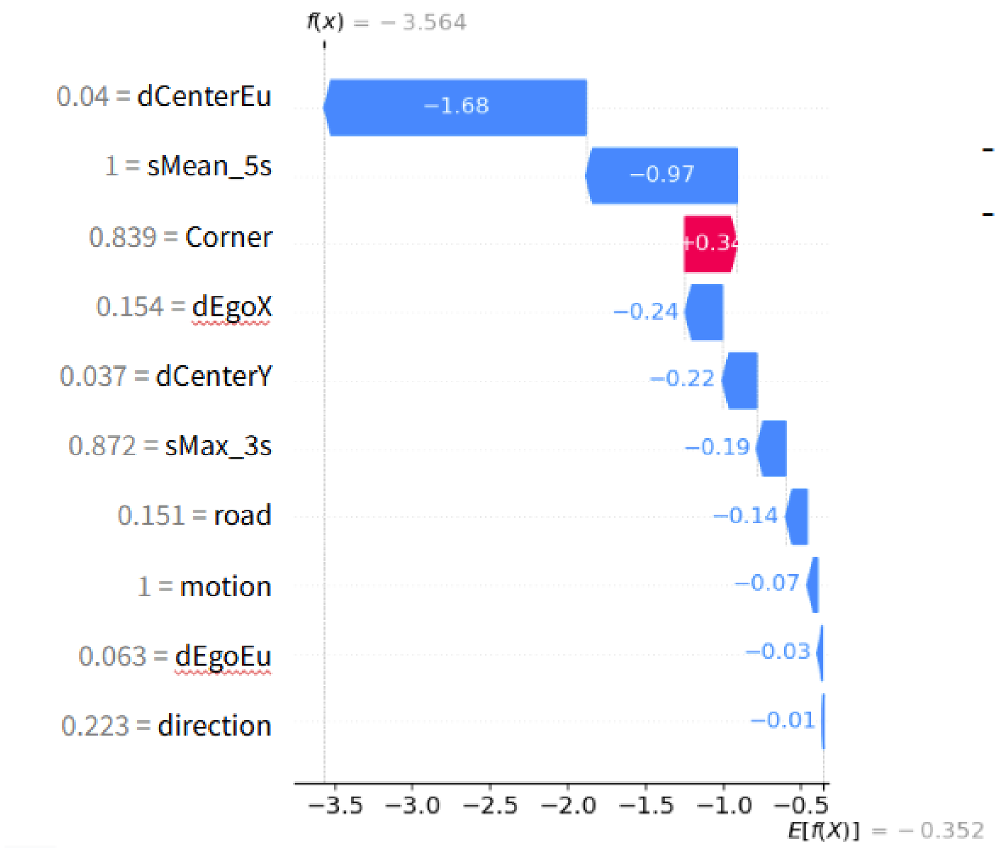


Figure 5.5: Local SHAP summary plot of low driver SA

Chapter 6

SUMMARY

The primary objective of this research is to develop real-time predictive models of driver SA, with a dual emphasis on performance and interpretability. The pursuit of this objective is guided by a methodical approach involving the following key steps:

1. **Identification of Comprehensive Predictors**
   The initial step entails identifying a comprehensive set of predictors crucial for driver SA prediction.

2. **Exploration of Relationships**
   Following predictor identification, the study delves into exploring the relationships between the three levels of SA (perception, comprehension, and projection) and the identified predictors.

3. **Development of SA Predictive Models**
   Subsequently, real-time predictive models for three levels of driver SA are developed based on the insights gained from the predictor-relationship exploration.

4. **Comparative Assessment**
   The final step involves a comparative assessment of the performance and interpretability of the developed models.

To achieve the first two objectives and gather essential data for subsequent phases, two driving simulator studies were conducted. These studies not only facilitated the identification of comprehensive predictors but also provided a platform for a detailed examination of the nuanced relationships between different levels of SA and the identified predictors. The data obtained from these studies serve as a foundational resource for the development of real-time predictive models and subsequent comparative analyses.

This work's contribution in relation to prior research is as follows:

1. Historically, driver SA prediction efforts predominantly leaned on singular factors such as either eye gaze or environmental considerations in isolation. This limited approach hindered the comprehensive development of models. In contrast, this thesis adopts a more inclusive methodology by concurrently considering these factors in the model development process.

2. Previous research primarily focused on the prediction of SA at level 1, often overlooking the significance of levels 2 and 3. In recognition of the importance of all three levels of SA, this thesis integrates predictions for levels 1, 2, and 3 of driver SA, providing a more holistic perspective.

3. Earlier works in the domain of driver SA predictive models frequently prioritized performance metrics while neglecting the essential aspect of interpretability. In contrast, this thesis places emphasis on interpretability, acknowledging its crucial role in advancing the field beyond mere performance evaluation.

BIBLIOGRAPHY

[1] Mohamed Abdel-Aty and Joanne Keller. Exploring the overall and specific crash severity levels at signalized intersections. *Accident Analysis & Prevention*, 37(3):417–425, 2005.

[2] Torbjörn Åkerstedt, Björn Peters, Anna Anund, and Göran Kecklund. Impaired alertness and performance driving home from the night shift: a driving simulator study. *Journal of sleep research*, 14(1):17–20, 2005.

[3] Yasir Ali, Anshuman Sharma, Md Mazharul Haque, Zuduo Zheng, and Mohammad Saifuzzaman. The impact of the connected environment on driving behavior and safety: A driving simulator study. *Accident Analysis & Prevention*, 144:105643, 2020.

[4] Theodore W Anderson. *The statistical analysis of time series*. John Wiley & Sons, 2011.

[5] Gianluca Antonini, Michel Bierlaire, and Mats Weber. Discrete choice models of pedestrian walking behavior. *Transportation Research Part B: Methodological*, 40(8):667–687, 2006.

[6] Jackie Ayoub, Feng Zhou, Shan Bao, and X Jessie Yang. From manual driving to automated driving: A review of 10 years of autoui. In *Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications*, pages 70–90, 2019.

[7] Gavin C Barr Jr, Kathleen E Kane, Robert D Barraco, Timarie Rayburg, Lauren Demers, Chadd K Kraus, Marna Rayl Greenberg, Valerie A Rupp, Kimberly M Hamilton, and Bryan G Kane. Gender differences in perceptions and self-reported driving behaviors among teenagers. *The Journal of emergency medicine*, 48(3):366–370, 2015.

[8] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[9] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. R package version 1.1-29.

[10] Kristen E Beede and Steven J Kass. Engrossed in conversation: The impact of cell phones on simulated driving performance. *Accident Analysis & Prevention*, 38(2):415–421, 2006.

[11] Anika Boelhouwer, Arie Paul van den Beukel, Mascha C Voort, and Marieke H Martens. Determining infrastructure-and traffic factors that increase the perceived complexity of driving situations. *International Conference on Applied Human Factors and Ergonomics*, 1212(2):3–10, 2020.

[12] Karel A Brookhuis and Dick de Waard. Assessment of drivers' workload: performance and subjective and physiological indexes. In *Stress, Workload, and Fatigue*, pages 321–333. CRC press, 2000.

[13] Johnell O Brooks, Richard A Tyrrell, and Talissa A Frank. The effects of severe visual challenges on steering performance in visually healthy young drivers. *Optometry and Vision Science*, 82(8):689–697, 2005.

[14] Jeff K. Caird, Chelsea R. Willness, Piers Steel, and Chip Scialfa. A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, 40(4):1282–1293, 2008.

[15] Jeff K Caird, Chelsea R Willness, Piers Steel, and Chip Scialfa. A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, 40(4):1282–1293, 2008.

[16] Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & electrical engineering*, 40(1):16–28, 2014.

[17] Peter Chapman, Geoffrey Underwood, and Katharine Roberts. Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2):157–167, 2002.

[18] Bo-Chiuan Chen, Bi-Cheng Luan, and Kangwon Lee. Design of lane keeping system using adaptive model predictive control. In *2014 IEEE International Conference on Automation Science and Engineering (CASE)*, pages 922–926, Taipei, Taiwan, 2014. IEEE.

[19] On-Road Automated Driving (ORAD) Committee. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, apr 2021.

[20] John G Cragg. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, 39:829–844, 1971.

[21] Lisa A D'Ambrosio, Laura KM Donorfio, Joseph F Coughlin, Maureen Mohyde, and Joachim Meyer. Gender differences in self-regulation patterns and attitudes toward driving among older adults. *Journal of women & aging*, 20(3-4):265–282, 2008.

[22] Dick De Waard, Anje Kruizinga, and Karel A Brookhuis. The consequences of an increase in heavy goods vehicles for passenger car drivers' mental workload and behaviour: a simulator study. *Accident analysis & prevention*, 40(2):818–828, 2008.

[23] Joost CF De Winter, Riender Happee, Marieke H Martens, and Neville A Stanton. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27:196–217, 2014.

[24] David M DeJoy. An examination of gender differences in traffic accident risk perception. *Accident Analysis & Prevention*, 24(3):237–246, 1992.

[25] T. A. Dingus, S. G. Klauer, V. L. Neale, A. Petersen, S. E. Lee, J. Sudweeks, M. a. Perez, J. Hankey, D. Ramsey, S. Gupta, C. Bucher, Z. R. Doerzaph, J. Jermeland, and R.R. Knipling. *The 100-Car naturalistic driving study phase II – results of the 100-Car field experiment*. Department of Transportation, 2006.

[26] Na Du, Feng Zhou, Elizabeth M Pulver, Dawn M Tilbury, Lionel P Robert, Anuj K Pradhan, and X Jessie Yang. Predicting driver takeover performance in conditionally automated driving. *Accident Analysis & Prevention*, 148:105748, 2020.

[27] Francis T Durso, Carla A Hackworth, Todd R Truitt, Jerry Crutchfield, Danko Nikolic, and Carol A Manning. Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Quarterly*, 6(1):1–20, 1998.

[28] Jessica Edquist. *The effects of visual clutter on driving performance*. PhD thesis, Monash University, 2008.

[29] Mica R Endsley. Situation awareness global assessment technique (SAGAT). In *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, pages 789–795, OH, USA, 1988. IEEE.

[30] Mica R Endsley. Measurement of situation awareness in dynamic systems. *Human Factors*, 37(1):65–84, 1995.

[31] Mica R Endsley. Toward a theory of situation awareness in dynamic systems. *Human factors*, 37(1):32–64, 1995.

[32] Mica R Endsley. Direct measurement of situation awareness: Validity and use of SAGAT. *Situation awareness analysis and measurement*, 10:147–173, 2000.

[33] Mica R Endsley. Final reflections: Situation awareness models and measures. *Journal of Cognitive Engineering and Decision Making*, 9(1):101–111, 2015.

[34] Mica R Endsley. Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*, 9(1):4–32, 2015.

[35] Mica R Endsley. Direct measurement of situation awareness: Validity and use of sagat. In *Situational awareness*, pages 129–156. Routledge, 2017.

[36] Mica R Endsley. The divergence of objective and subjective situation awareness: A meta-analysis. *Journal of cognitive engineering and decision making*, 14(1):34–53, 2020.

[37] Mica R Endsley. Situation awareness. In *Handbook of human factors and ergonomics*, pages 434–455. Wiley Online Library, 2021.

[38] Mica R Endsley. A systematic review and meta-analysis of direct objective measures of situation awareness: a comparison of SAGAT and SPAM. *Human Factors*, 63(1):124–150, 2021.

[39] Mica R Endsley and W Jones. Situation awareness. *The Oxford handbook of cognitive engineering*, 1:88–108, 2013.

[40] Mica R. Endsley, Stephen J. Selcon, Thomas D. Hardiman, and Darryl G. Croft. A comparative analysis of SAGAT and SART for evaluations of situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 42(1):82–86, 1998.

[41] Berthold Färber. Communication and communication problems between autonomous vehicles and human drivers. *Autonomous driving: Technical, legal and social aspects*, pages 125–144, 2016.

[42] Alex A Freitas. Comprehensible classification models: a position paper. *ACM SIGKDD explorations newsletter*, 15(1):1–10, 2014.

[43] Mitsuki Fujino, Jieun Lee, Toshiaki Hirano, Yuichi Saito, and Makoto Itoh. Comparison of sagat and spam for seeking effective way to evaluate situation awareness and workload during air traffic control task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 64(1):1836–1840, 2020.

[44] Christian Gold, Moritz Körber, David Lechner, and Klaus Bengler. Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. *Human factors*, 58(4):642–652, 2016.

[45] Paul Green. Why driving performance measures are sometimes not accurate (and methods to check accuracy). In *Driving Assessment Conference*, volume 4. University of Iowa, 2007.

[46] Paul E Green, Takahiro Wada, Jessica Oberholtzer, Paul A Green, Jason Schweitzer, and Hong Eoh. How do distracted and normal driving differ: An analysis of the acas naturalistic driving data. Technical report, University of Michigan. Transportation Research Institute, 2007.

[47] Jeff Greenberg, Bruce Artz, and Larry Cathey. The effect of lateral motion cues during simulated driving. *Proceedings of DSC North America*, pages 8–10, 2003.

[48] David Gunning and David Aha. Darpa's explainable artificial intelligence (xai) program. *AI magazine*, 40(2):44–58, 2019.

[49] Maya Gupta, Andrew Cotter, Jan Pfeifer, Konstantin Voevodski, Kevin Canini, Alexander Mangylov, Wojciech Moczydlowski, and Alexander Van Esbroeck. Monotonic calibrated interpolated look-up tables. *The Journal of Machine Learning Research*, 17(1):3790–3836, 2016.

[50] Robert S Gutzwiller and Benjamin A Clegg. The role of working memory in levels of situation awareness. *Journal of Cognitive Engineering and Decision Making*, 7(2):141–154, 2013.

[51] DE Haigney, RG Taylor, and SJ Westerman. Concurrent mobile (cellular) phone use and driving performance: task demand characteristics and compensatory processes. *Transportation Research Part F: Traffic Psychology and Behaviour*, 3(3):113–121, 2000.

[52] Xueqin Hao, Zhiguo Wang, Fan Yang, Ying Wang, Yanru Guo, and Kan Zhang. The effect of traffic on situation awareness and mental workload: Simulator-based study. In *International Conference on Engineering Psychology and Cognitive Ergonomics*, pages 288–296, Berlin, Heidelberg, 2007. Springer.

[53] Md Mazharul Haque and Simon Washington. The impact of mobile phone distraction on the braking behaviour of young drivers: a hazard-based duration model. *Transportation research part C: emerging technologies*, 50:13–27, 2015.

[54] Joanne L Harbluk, Y Ian Noy, Patricia L Trbovich, and Moshe Eizenman. An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention*, 39(2):372–379, 2007.

[55] Geoffrey Ho, Charles T Scialfa, Jeff K Caird, and Trevor Graw. Visual search for traffic signs: The effects of clutter, luminance, and aging. *Human factors*, 43(2):194–207, 2001.

[56] Markus Hofbauer, Christopher B Kuhn, Jiaming Meng, Goran Petrovic, and Eckehard Steinbach. Multi-view region of interest prediction for autonomous driving using semi-supervised labeling. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6. IEEE, 2020.

[57] William J Horrey and Christopher D Wickens. Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human factors*, 48(1):196–205, 2006.

[58] Johan Huysmans, Karel Dejaeger, Christophe Mues, Jan Vanthienen, and Bart Baesens. An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. *Decision Support Systems*, 51(1):141–154, 2011.

[59] Kentaro Iio, Xiaoyu Guo, and Dominique Lord. Examining driver distraction in the context of driving speed: An observational study using disruptive technology and naturalistic data. *Accident Analysis & Prevention*, 153:105983, 2021.

[60] Lisheng Jin, Qingning Niu, Haijing Hou, Huacai Xian, Yali Wang, Dongdong Shi, et al. Driver cognitive distraction detection using driving performance measures. *Discrete Dynamics in Nature and Society*, 2012, 2012.

[61] Marcel A Just and Patricia A Carpenter. A theory of reading: from eye fixations to comprehension. *Psychological review*, 87(4):329, 1980.

[62] David Kaber, Yu Zhang, Sangeun Jin, Prithima Mosaly, and Megan Garner. Effects of hazard exposure and roadway complexity on young and older driver situation awareness and performance. *Transportation research part F: traffic psychology and behaviour*, 15(5):600–611, 2012.

[63] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[64] Lisa Keay, Srichand Jasti, Beatriz Munoz, Kathleen A Turano, Cynthia A Munro, Donald D Duncan, Kevin Baldwin, Karen J Bandeen-Roche, Emily W Gower, and Sheila K West. Urban and rural differences in older drivers' failure to stop at stop signs. *Accident Analysis & Prevention*, 41(5):995–1000, 2009.

[65] Hyungil Kim, Joseph L Gabbard, Sujitha Martin, Ashish Tawari, and Teruhisa Misu. Toward prediction of driver awareness of automotive hazards: Driving-video-based simulation approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63:2099–2103, 2019.

[66] Hyungil Kim, Sujitha Martin, Ashish Tawari, Teruhisa Misu, and Joseph L Gabbard. Toward real-time estimation of driver situation awareness: An eye-tracking approach based on moving objects of interest. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1035–1041, Las Vegas, USA, 2020. IEEE.

[67] Max Kuhn. *caret: Classification and Regression Training*, 2022. R package version 6.0-93.

[68] Miltos Kyriakidis, Joost CF de Winter, Neville Stanton, Thierry Bellet, Bart van Arem, Karel Brookhuis, Marieke H Martens, Klaus Bengler, Jan Andersson, Natasha Merat, et al. A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science*, 20(3):223–249, 2019.

[69] Neil D Lerner. *Additional investigations on driver information overload*, volume 36. Transportation Research Board, Washington, DC, 2003.

[70] Yang Li, Keqiang Li, Yang Zheng, Bernhard Morys, Shuyue Pan, and Jianqiang Wang. Threat assessment techniques in intelligent vehicles: A comparative survey. *IEEE Intelligent Transportation Systems Magazine*, 13(4):71–91, 2020.

[71] Nade Liang, Jing Yang, Denny Yu, Kwaku O Prakah-Asante, Reates Curry, Mike Blommer, Radhakrishnan Swaminathan, and Brandon J Pitts. Using eye-tracking to investigate the effects of pre-takeover visual engagement on situation awareness during automated driving. *Accident Analysis & Prevention*, 157:106143, 2021.

[72] Yulan Liang and John D Lee. Combining cognitive and visual distraction: Less than the sum of its parts. *Accident Analysis & Prevention*, 42(3):881–890, 2010.

[73] Yin Lou, Rich Caruana, Johannes Gehrke, and Giles Hooker. Accurate intelligible models with pairwise interactions. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 623–631, 2013.

[74] WJR Louwerse and SP Hoogendoorn. Adas safety impacts on rural and urban highways. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 887–890. IEEE, 2004.

[75] Guangquan Lu, Junda Zhai, Penghui Li, Facheng Chen, and Liming Liang. Measuring drivers' takeover performance in varying levels of automation: Considering the influence of cognitive secondary task. *Transportation Research Part F: Traffic Psychology and Behaviour*, 82:96–110, 2021.

[76] Zhenji Lu, Riender Happee, and Joost CF de Winter. Take over! a video-clip study measuring attention, situation awareness, and decision-making in the face of an impending hazard. *Transportation Research Part F: Traffic Psychology and Behaviour*, 72:211–225, 2020.

[77] Nils Lubbe. Brake reactions of distracted drivers to pedestrian forward collision warning systems. *Journal of safety research*, 61:23–32, 2017.

[78] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. *Nature machine intelligence*, 2(1):56–67, 2020.

[79] Nengchao Lyu, Lian Xie, Chaozhong Wu, Qiang Fu, and Chao Deng. Driver's cognitive workload and driving performance under traffic sign information exposure in complex environments: A case study of the highways in china. *International journal of environmental research and public health*, 14(2):203, 2017.

[80] Andrew K Mackenzie and Julie M Harris. A link between attentional function, effective eye movements, and driving ability. *Journal of experimental psychology: human perception and performance*, 43(2):381, 2017.

[81] Dominique Makowski, Mattan S. Ben-Shachar, Indrajeet Patil, and Daniel Ludecke. Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption, 2021.

[82] Michael P Manser and Peter A Hancock. The influence of perceptual speed regulation on speed perception, choice, and control: Tunnel wall characteristics and influences. *Accident Analysis & Prevention*, 39(1):69–78, 2007.

[83] Wilson E Marcílio and Danilo M Eler. From explanations to feature selection: assessing shap values as feature selection mechanism. In *2020 33rd SIBGRAPI conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 340–347. Ieee, 2020.

[84] Laila M Martinussen, Timo Lajunen, Mette Møller, and Türker Özkan. Short and user-friendly: The development and validation of the mini-dbq. *Accident Analysis & Prevention*, 50:1259–1265, 2013.

[85] Michael L Matthews, David J Bryant, Robert D G Webb, Joanne L Harbluk, M L Matthews, D J Bryant, and R D G Webb. Model for situation awareness and driving application to analysis and research for intelligent transportation systems, 2001.

[86] S Mayer. Imagefluency: Image statistics based on processing fluency, 2021. R package version 0.2.3.

[87] Daniel V McGehee, John D Lee, Matthew Rizzo, Jeffrey Dawson, and Kirk Bateman. Quantitative analysis of steering adaptation on a high performance fixed-base driving simulator. *Transportation Research Part F: Traffic Psychology and Behaviour*, 7(3):181–196, 2004.

[88] Yongyi Min and Alan Agresti. Modeling nonnegative data with clumping at zero: a survey. *Journal of the Iranian Statistical Society (JIRSS)*, 1:7–33, 2002.

[89] Yongyi Min and Alan Agresti. Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling*, 5(1):1–19, 2005.

[90] Lisa J Molnar. Age-related differences in driver behavior associated with automated vehicles and the transfer of control between automated and manual control: A simulator evaluation. Technical report, University of Michigan, Ann Arbor, Transportation Research Institute, 2017.

[91] Kristin Moore and Leo Gugerty. Development of a novel measure of situation awareness: The case for eye movement analysis. In *Proceedings of the human factors and ergonomics society annual meeting*, number 19 in 1, pages 1650–1654. SAGE Publications Sage CA: Los Angeles, CA, 2010.

[92] Abdallah Moujahid, Mounir ElAraki Tantaoui, Manolo Dulva Hina, Assia Soukane, Andrea Ortalda, Ahmed ElKhadimi, and Amar Ramdane-Cherif. Machine learning techniques in adas: A review. In *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, pages 235–242, Paris, France, 2018. IEEE.

[93] John Mullahy. Specification and testing of some modified count data models. *Journal of Econometrics*, 33(3):341–365, 1986.

[94] Nadia Mullen, Judith Charlton, Anna Devlin, and Michel Bedard. Simulator validity: Behaviours observed on the simulator and on the road. In *Handbook of driving simulation for engineering, medicine and psychology*, pages 1–18. CRC Press, 2011.

[95] Thanh Nguyen, Chee Peng Lim, Ngoc Duy Nguyen, Lee Gordon-Brown, and Saeid Nahavandi. A review of situation awareness assessment approaches in aviation environments. *IEEE Systems Journal*, 13(3):3590–3603, 2019.

[96] Panagiotis Papantoniou. Structural equation model analysis for the evaluation of overall driving performance: A driving simulator study focusing on driver distraction. *Traffic Injury Prevention*, 19(3):317–325, 2018.

[97] Panagiotis Papantoniou, Eleonora Papadimitriou, and George Yannis. Review of driving performance parameters critical for distracted driving research. *Transportation Research Procedia*, 25:1796–1805, 2017. World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016.

[98] Panagiotis Papantoniou, Eleonora Papadimitriou, and George Yannis. Review of driving performance parameters critical for distracted driving research. *Transportation research procedia*, 25:1796–1805, 2017.

[99] Sami Park, Yilun Xing, Kumar Akash, Teruhisa Misu, and Linda Ng Boyle. The impact of environmental complexity on drivers' situation awareness. In *Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 131–138, 2022.

[100] Sami Park, Yilun Xing, Kumar Akash, Teruhisa Misu, Shashank Mehrotra, and Linda Ng Boyle. The impact of pedestrian interactions at intersections on the three levels of drivers' situation awareness. *Transportation Research Part F: Traffic Psychology and Behaviour*, 107:167–180, 2024.

[101] Christopher JD Patten, Albert Kircher, Joakim Östlund, Lena Nilsson, and Ola Svenson. Driver experience and cognitive workload in different traffic environments. *Accident Analysis & Prevention*, 38(5):887–894, 2006.

[102] Jonas Radlmayr, Christian Gold, Lutz Lorenz, Mehdi Farid, and Klaus Bengler. How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 58, pages 2063–2067, Los Angeles, CA, 2014. Sage Publications Sage CA.

[103] Michael E Rakauskas, Leo J Gugerty, and Nicholas J Ward. Effects of naturalistic cell phone conversations on driving performance. *Journal of safety research*, 35(4):453–464, 2004.

[104] Vasili Ramanishka, Yi-Ting Chen, Teruhisa Misu, and Kate Saenko. Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7699–7707, Salt Lake City, UT, USA, 2018. IEEE.

[105] Thomas A Ranney, Joanne L Harbluk, and Y Ian Noy. Effects of voice technology on test track driving performance: Implications for driver distraction. *Human factors*, 47(2):439–454, 2005.

[106] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. Understanding pedestrian behavior in complex traffic scenes. *IEEE Transactions on Intelligent Vehicles*, 3(1):61–70, 2017.

[107] James Reason, Antony Manstead, Stephen Stradling, James Baxter, and Karen Campbell. Errors and violations on the roads: a real distinction? *Ergonomics*, 33(10-11):1315–1332, 1990.

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

[109] Umair Rehman, Cao Shi, and MacGregor Carolyn. Using an integrated cognitive architecture to model the effect of environmental complexity on drivers' situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1):812–816, 2019.

[110] Nancy Rhodes and Kelly Pivik. Age and gender differences in risky driving: The roles of positive affect and risk perception. *Accident Analysis & Prevention*, 43(3):923–931, 2011.

[111] Michele Rucci, Paul V McGraw, and Richard J Krauzlis. Fixational eye movements and perception. *Vision research*, 118:1–4, 2016.

[112] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *AI magazine*, 11 2018.

[113] Stefan Rüping. Learning interpretable models. *AI magazine*, 2006.

[114] SAE. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles, 2021.

[115] Bobbie Seppelt and C Wickens. In-vehicle tasks: Effects of modality, driving relevance. Technical report, and redundancy. Technical Report AHFD-03-16 & GM-03-2, Aviation Human . . . , 2003.

[116] F Atiyya Shaw, Aaron T Greenwood, JongIn Bae, Gregory M Corso, Michael O Rodgers, and Michael P Hunter. Effects of roadway factors and demographic characteristics on drivers' perceived complexity of simulated roadway videos. *Transportation Letters*, 11(10):589–598, 2019.

[117] Jaeyoung Shin. Feasibility of local interpretable model-agnostic explanations (lime) algorithm as an effective and interpretable feature selection method: comparative fnirs study. *Biomedical Engineering Letters*, 13(4):689–703, 2023.

[118] Santokh Singh. Critical reasons for crashes investigated in the national motor vehicle crash causation survey. Technical report, NHTSA's National Center for Statistics and Analysis, 2015.

[119] Wenjie Song, Yi Yang, Mengyin Fu, Yujun Li, and Meiling Wang. Lane detection and classification for forward collision warning system based on stereo vision. *IEEE Sensors Journal*, 18(12):5151–5163, 2018.

[120] Rebecca Spicer, Amin Vahabaghaie, Dennis Murakhovsky, George Bahouth, Becca Drayer, and Schuyler St Lawrence. Effectiveness of advanced driver assistance systems in preventing system-relevant crashes. *SAE International Journal of Advances and Current Practices in Mobility*, 3(2021-01-0869):1697–1701, 2021.

[121] Petre Stoica and Yngve Selen. Model-order selection: a review of information criterion rules. *IEEE Signal Processing Magazine*, 21(4):36–47, 2004.

[122] Foghor Tanshi and Dirk Söffker. Modeling of takeover variables with respect to driver situation awareness and workload for intelligent driver assistance. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, pages 1667–1672. IEEE, 2019.

[123] Evona Teh, Samantha Jamson, Oliver Carsten, and Hamish Jamson. Temporal fluctuations in driving demand: The effect of traffic complexity on subjective measures of workload and driving performance. *Transportation research part F: traffic psychology and behaviour*, 22:207–217, 2014.

[124] Duy Tran, Ha Manh Do, Weihua Sheng, He Bai, and Girish Chowdhary. Real-time detection of distracted driving based on deep learning. *IET Intelligent Transport Systems*, 12(10):1210–1219, 2018.

[125] University of Iowa. National advanced driving simulator (nads) minisim, 2011. Accessed August 11, 2023. https://www.nads-sc.uiowa.edu/minisim/.

[126] B Venkatesh and J Anuradha. A review of feature selection and its methods. *Cybernetics and information technologies*, 19(1):3–26, 2019.

[127] Guy H Walker, Neville A Stanton, and Mark S Young. Feedback and driver situation awareness (sa): A comparison of sa measures and contexts. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(4):282–299, 2008.

[128] Julia Werneke and Mark Vollrath. What does the driver look at? the influence of intersection characteristics on attention allocation and driving behavior. *Accident Analysis & Prevention*, 45:610–619, 2012.

[129] Michał Wierzchoń, Anna Anzulewicz, Justyna Hobot, Borysław Paulewicz, and Jérôme Sackur. In search of the optimal measure of awareness: Discrete or continuous? *Consciousness and cognition*, 75:102798, 2019.

[130] Julia L Wright, Jessie YC Chen, and Shan G Lakhmani. Agent transparency and reliability in human–robot interaction: The influence on user confidence and perceived reliability. *IEEE Transactions on Human-Machine Systems*, 50(3):254–263, 2019.

[131] Tong Wu, Enna Sachdeva, Kumar Akash, Xingwei Wu, Teruhisa Misu, and Jorge Ortiz. Toward an adaptive situational awareness support system for urban driving. In *2022 IEEE Intelligent Vehicles Symposium (IV)*, pages 1073–1080. IEEE, 2022.

[132] Yanbin Wu, Ken Kihara, Yuji Takeda, Toshihisa Sato, Motoyuki Akamatsu, Satoshi Kitazaki, Koki Nakagawa, Kenta Yamada, Hiromitsu Oka, and Shougo Kameyama. Eye movements predict driver reaction time to takeover request in automated driving: A real-vehicle study. *Transportation Research part F: Traffic Psychology and Behaviour*, 81:355–363, 2021.

[133] Rachael A Wynne, Vanessa Beanland, and Paul M Salmon. Systematic review of driving simulator validation studies. *Safety Science*, 117:138–151, 2019.

[134] Hui Xi and Young-Jun Son. Two-level modeling framework for pedestrian route choice and walking behaviors. *Simulation Modelling Practice and Theory*, 22:28–46, 2012.

[135] Yilun Xing, Sami Park, Kumar Akash, Teruhisa Misu, Shashank Mehrotra, and Linda Ng Boyle. The impact of environmental features on drivers' situation awareness using real-world driving scenarios. *International Journal of Human–Computer Interaction*, 39(16):3203–3212, 2023.

[136] Yilun Xing, Sami Park, Kumar Akash, Xingwei Wu, Teruhisa Misu, and Linda Ng Boyle. Investigating the impact of context and environment on driver's situation awareness. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 66(1):335–339, 2022.

[137] Jing Yang, Nade Liang, Brandon J. Pitts, Kwaku O. Prakah-Asante, Reates Curry, Mike Blommer, Radhakrishnan Swaminathan, and Denny Yu. Multimodal sensing and computational intelligence for situation awareness classification in autonomous driving. *IEEE Transactions on Human-Machine Systems*, 53(2):270–281, 2023.

[138] George Yannis, Eleonora Papadimitriou, Xenia Karekla, and Efrosyni Kontodima. Mobile phone use by young drivers: effects on traffic speed and headways. *Transportation planning and technology*, 33(4):385–394, 2010.

[139] Kristie Young, Michael Regan, and Mike Hammer. Driver distraction: A review of the literature. *Distracted driving*, 2007:379–405, 2007.

[140] Kristie L Young, Paul M Salmon, and Miranda Cornelissen. Missing links? the effects of distraction on driver situation awareness. *Safety science*, 56:36–43, 2013.

[141] Kathrin Zeeb, Axel Buchner, and Michael Schrauf. Is take-over time all that matters? the impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis & Prevention*, 92:230–239, 2016.

[142] Ting Zhang, Jing Yang, Nade Liang, Brandon J Pitts, Kwaku O Prakah-Asante, Reates Curry, Bradley S Duerstock, Juan P Wachs, and Denny Yu. Physiological measurements of situation awareness: a systematic review. *Human factors*, page 0018720820969071, 2020.

[143] Hongtao Zheng, Tongtong Zhou, Ting Han, Shuo Li, and Cong Yu. An interpretable prediction framework for multi-class situational awareness in conditionally automated driving. *Advanced Engineering Informatics*, 62:102683, 2024.

[144] Feng Zhou, X Jessie Yang, and Joost CF De Winter. Using eye-tracking data to predict situation awareness in real time during takeover transitions in conditionally automated driving. *IEEE Transactions on Intelligent Transportaion Systems*, 23(3):2284–2295, 2021.

[145] Haibei Zhu, Teruhisa Misu, Sujitha Martin, Xingwei Wu, and Kumar Akash. Improving driver situation awareness prediction using human visual sensory and memory mechanism. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6210–6216, Prague, Czech Republic, 2021. IEEE.

[146] Siying Zhu. Analyse vehicle–pedestrian crash severity at intersection with data mining techniques. *International journal of crashworthiness*, 27(5):1374–1382, 2022.

Appendix A

PRE-EXPERIMENT QUESTIONNAIRE OF STUDY 1 AND 2

As part of this study, it is useful to collect information describing each participant. The following questions ask about your basic demographic information, health, driving frequency and patterns. Please read each question carefully. If something is unclear, ask the research assistant for help. Your participation is voluntary, and you may skip any questions that you do not want to answer.

A.1 Demographic Information

1) What is your birth date?

2) What is your gender?

3) Do you wear contact lenses?

4) Do you wear glasses when driving?

5) What is your highest level of education?

A.2 Driving History

6) How old were you when you started to drive (including driving before license)?

7) At what age did you get your first driver's license (not just a permit)?

8) How often do you drive?

9) How many miles did you drive last week?

10) a. How many moving violations have you had in the last 3 years:

b. How many of these moving violations were in the past year:

11) a. How many vehicular crashes have you been involved in during the last 3 years:

b. How many of these crashes were considered 'your fault':

12) a. How many vehicular crashes have you been involved in during the past year:

b. How many of these crashes were considered 'your fault':

A.3 Vision and Hearing

13) What type of prescription glasses or contact lenses will you be wearing for today's study?

14) What type of vision problem do you have?

Appendix B

DRIVER BEHAVIOR QUESTIONNAIRE OF STUDY 2

The Driver Behavior Questionnaire (DBQ) is one of the most widely used instruments for measuring self-reported driving behaviors. In this study, a short and user-friendly version that consists of 9 questions is used [84].

B.1 In the past months while driving, how often did you...

1. Drive especially close or "flash" the car in front as a signal for that driver to go faster or get out of your way
2. Deliberately disregard the speed limits late at night or very early in the morning
3. Get involved in unofficial "races" with other car drivers
4. Turn right on to a main road into the path of an oncoming vehicle that you had not seen, or whose speed you had misjudged
5. Fail to notice, because lost in thought or distracted, someone waiting at a zebra crossing, or that a pelican crossing light has just turned red
6. Misjudge the speed of a moving vehicle when overtaking
7. Forget where you left your car in a multi-level car park
8. Get into the wrong lane at a roundabout or approach a road junction
9. Fail to read the signs correctly, and exit from a roundabout on the wrong road

Appendix C

SATISFACTION AND NASA TLX QUESTIONNAIRE OF STUDY 2

C.1 Satisfaction

1. How many intersections were you able to safely interact with pedestrians out of 6 event intersections?
2. Were there any intersections where you may have endangered a pedestrian? If yes, describe which intersection it was.
3. During the drive overall, how satisfied were you with your interactions with pedestrians?

C.2 NASA TLX

1. How mentally demanding was the task?
2. How physically demanding was the task?
3. How hurried or rushed was the pace of the task?
4. How successful were you in accomplishing what you were asked to do?
5. How hard did you have to work to accomplish your level of performance?
6. How insecure, discouraged, irritated, stressed, and annoyed were you?