Measuring and Predicting Driver Situation Awareness

Yilun Xing

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

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Linda Ng Boyle, Chair

Youngjun Choe

Ji-Eun Kim

Program Authorized to Offer Degree:
Industrial & Systems Engineering

University of Washington

Abstract

Measuring and Predicting Driver Situation Awareness

Yilun Xing

Chair of the Supervisory Committee:
Linda Ng Boyle
Industrial & Systems Engineering, Civil & Environmental Engineering

Situation awareness (SA) encompasses the perception, comprehension, and projection of elements within a given situation, aligning with the three levels of SA. It plays a critical role in driver safety, as drivers must continuously maintain awareness of dynamic road conditions, significantly influencing traffic crash rates associated with human error. Advanced driver assistance systems (ADAS) are designed to enhance driving performance. However, overly complex or ambiguous information may overwhelm drivers' limited SA capacity. To address this challenge, ADAS models should consider the operator's perception and understanding of the environment to tailor assistance effectively. This dissertation proposes methods to measure and predict driver SA, focusing on objects of interest within the scene. Initially, an experimental approach utilizing real-world driving videos and a web-based touch recorder was introduced to capture driver SA, which demonstrates efficacy in capturing various levels of driver SA. Furthermore, it was observed that drivers often rely on the memory of element trajectories to understand or predict the location of OOIs. Subsequent analysis identified the impacts of environmental features, object characteristics, and driver demographics on driver SA. Incorporating these feature groups into predictive models proved reasonable, with environmental factors such as the number of objects in the scene, scene visual complexity and roadway type, object features such as object size and type, and driver demographics such as gender showing significant impacts on driver SA. Next, gaze-point-based and visual sensory-dependent features were processed from the eye-tracking data. Predictive models

incorporating different feature groups including environmental, object, and driver features, and those extracted from the eye-tracking data were fitted and compared. Two phases of SA were distinguished: object localization and recognition. Binary classification models were developed and rigorously evaluated for each phase. Recognizing the impracticality of drivers wearing eye-trackers during daily driving, an alternative for eye-tracking data, which utilizes computer vision to estimate visual attention from forward-view driving videos was proposed. Gaze-related features extracted from these videos demonstrated comparable performance to those from eye-tracking data, suggesting their viability for predicting driver SA. This insight could inform the design of ADAS systems, enabling low-cost selective assistance to drivers and ultimately enhancing driver safety.

TABLE OF CONTENTS

Page

LIST OF FIGURES

ACKNOWLEDGMENTS

I want to express my heartfelt thanks to everyone who played a part in completing this dissertation.

I'm especially grateful to my supervisor, Prof. Linda Ng Boyle, for her invaluable guidance, constant support, and helpful feedback throughout this research journey. Her expertise, encouragement, and dedication have been crucial in shaping this dissertation. I loved it so much every time she said "I like the idea, go for it!"

I also want to extend my sincere appreciation to my committee, Prof. Ji-Eun Kim, Prof. Youngjun Choe, and Prof. Andrea Stocco for their insightful comments, suggestions, and assistance in reviewing and refining this dissertation.

To my colleagues and peers, thank you for your friendship, encouragement, and engaging discussions that enriched my academic experience. Special thanks to Sami Park, David Prendez, Jiaxin Li, Grace Douglas, and Mayuree Binjolkar for making our time at the HFSM lab enjoyable and memorable. I will always miss the weekly HFSM lab coffee. Thank you all for being a part of my PhD life over the past couple of years. I also want to express my gratitude to the members of the HFES student chapter at UW. Prof. Ji-Eun Kim, Jiaxin Li, David Prendez, Cherin Lim, and Grace Douglas, thank you for your hard work and for the feeling of belonging.

This dissertation is based on a project in collaboration with the Honda Research Institute, and I am grateful to Teruhisa Misu, Kumar Akash, Xingwei Wu, and Shashank K Mehrotra for their guidance and collaboration. Special thanks to Huizhong Guo, a senior PhD who has graduated from the HFSM lab, for your collaboration at the very beginning of the study, for brainstorming with me even after you graduated, and for being a good friend.

The first three years of my PhD journey coincided with the COVID-19 pandemic. Before

I got to explore the new campus life here, quarantine started and everything became virtual. I feel so lucky to have been living with my roommate Xiaonan Sun. Thanks to her for always being there for me, for all the chats, support, and care, and for feeling like family. I will always remember the time we spent together.

Lastly, a big thank you to my family, especially my fiancé Zeyu Wei, for their unwavering love, support, and encouragement. You've been my rock throughout this journey.

Thank you to everyone who helped me during the journey, no matter how small. Your support has been invaluable and deeply appreciated.

DEDICATION

To My Parents And To My Love Zeyu

To my grandfather, Anchen Xing, who was a professor. Unfortunately, due to the pandemic, I didn't have the opportunity to meet him since I came to UW for this PhD program. He passed away before I could return earlier this year. I've always believed he was the one most eager to witness my graduation.

Chapter 1

INTRODUCTION

Situational awareness (SA) is defined as three levels: "the perception of environmental elements and events with respect to time or space, the comprehension of their meaning, and the projection of their future status" [31]. For instance, in a scenario involving a pedestrian on a sidewalk, SA entails recognizing the pedestrian's presence, understanding their current actions (such as standing or preparing to cross), and predicting their forthcoming movements (including trajectory and speed). These levels of SA represent progressive stages rather than a linear sequence [33]. The SA process can follow a strictly data-driven progression from perception to comprehension to projection. Alternatively, it may be influenced by individuals' goals and current understanding, prompting them to seek additional information to confirm or challenge their assessments.

SA holds significant importance for Advanced Driver Assistance Systems (ADAS), which rely on understanding driver perception to offer selective assistance. ADAS encompasses in-vehicle technologies like forward collision warning systems and lane-keeping assist, all aimed at assisting drivers in appropriately attending to the roadway when needed [95, 15]. Effective assistance or intervention from these systems hinges on enhancing drivers' awareness of their surroundings. Present systems furnish drivers with feedback based on vehicle data (e.g., braking, acceleration, steering) and environmental conditions. A key challenge for human operators lies in promptly interpreting these displays to take necessary action. As vehicles move towards greater autonomy, the exchange of information between controllers—be it the driver or the vehicle itself—must become more attuned to each other's perceptions and future intentions. This becomes particularly critical for Society of Automotive Engineers (SAE) levels 2 and 3 automation, where control is expected to transition between driver and vehicle [90].

Even as driving tasks evolve towards automation, the role of the human driver remains

substantial. Driving involves intricate cognitive processes, including visually scanning the environment for potential threats, predicting unseen dangers, making decisions, and executing responses [36]. Workload and situation awareness are identified as pivotal factors predictive of performance and safety [20].

Current in-vehicle systems are not often necessarily integrated, resulting in a barrage of simultaneous, non-prioritized alerts directed at drivers. This can elevate driver workload and hinder their ability to direct attention effectively in critical situations [10, 81].

By incorporating SA considerations, ADAS can be tailored to issue alerts pertaining specifically to objects requiring driver attention. Furthermore, integrating SA information can accommodate fluctuations in driver SA levels—perception, comprehension, and projection—providing personalized feedback and fostering a deeper understanding of the safety context.

The primary goal of this study is to measure driver SA and develop an integrated predictive model for driver SA applicable in real-time scenarios. The rest of this proposal is organized as follows. Chapter 2 provides an overview of relevant literature and identifies gaps in previous research. Chapter 3 outlines the research questions. Chapter 4 delineates the experimental methodology and procedures for acquiring driver SA and eye-tracking data. In Chapter 5, preliminary analyses showcase factors impacting driver SA, including environmental, object, and driver features. Chapter 6 delves into the processing of eye-tracking data and the evaluation of predictive models integrating features derived from this data. Chapter 7 introduces an alternative approach utilizing computer vision techniques to estimate drivers' visual attention from forward-view driving videos. Predictive models incorporating these estimated eye-gaze-related features are assessed against those relying on actual eye-tracking data. Finally, Chapter 8 serves as a general conclusion, highlighting contributions, acknowledging limitations, and suggesting potential paths for future research.

Chapter 2

RELATED WORK

2.1 Measurement of SA

The relations between various situation awareness measures are depicted in Figure 2.1. The four columns represent the different stages of human information processing in a driving task: information assessment, situation awareness generating, decision-making, and action. Measures taken during the situation awareness generation stage are considered direct, while those taken during the other stages are considered indirect.

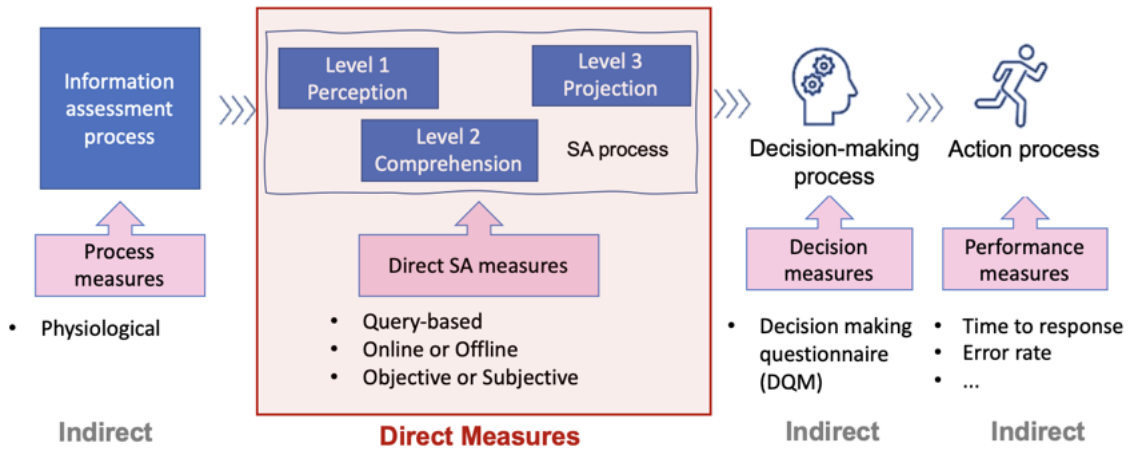


Figure 2.1: Situation awareness measures at different stages of human information processing in driving task

Process measures include those physiological measures recorded during the information assessment process such as eye-tracking measures, cardiovascular measures, and electroencephalography measures [107]. Both theoretical and methodological developments of the physiological measures have not reached the level of maturity as the concept of mental workload. Previous studies suggest that these physiological sensing methods may also be

capable of inferring SA. The most popular physiological measure, eye-tracking, provides a real-time, continuous measure of SA. Though it only indicates level 1 SA (exceptions include cases of "look but not see") without assessing the information through other senses such as auditory [27], correlations between conscious aspects of eye movement measures and direct SA scores have been observed [107]. Evidence for cardiovascular predictors of SA was mixed. Besides, SA has not been hypothesized to be associated with specific cortical regions or other physiological responses [99]: results from three electroencephalography (EEG) studies discovered that EEG was sensitive to changes in SA. Process measures also include communication measures (communication content, flow) which are mainly useful in team settings [27, 99].

There exist direct measures during the SA generating process, contradictory to indirect measures – the rest of the measures are categorized as indirect. The most common direct assessment techniques are query methods, which ask drivers about their awareness of the current situation. These methods can be divided into online and offline, based on whether the task or scenario is disrupted when the queries are given. The accuracy of the driver's responses to the queries determines their situation awareness score. Furthermore, direct measures can also be classified as objective and subjective, depending on whether the queries are based on factual information (e.g., whether the driver has observed something) or on the driver's subjective feelings (e.g., the confidence level of their responses). The widely used direct measure is the Situation Awareness Global Assessment Technique (SAGAT) [35], which is an offline and objective measure of situation awareness. Unlike online methods, SAGAT pauses the ongoing task to allow for more thorough probes, and it provides objective measures based on factual information, making it suitable for training predictive models. Thus, SAGAT was selected for use in this study.

Decision-making and performance are generated based on the situation awareness process, and the impact of the combination of all three levels. Park, Yoon, & Lee [79] used SA-related dimensions in the Decision Making Questionnaire (DMQ) to measure the decision-making style of participants in order to reflect a good SA. Dimensions selected include control (good SA $\rightarrow$ better control of the situation $\rightarrow$ efficient decision), thoroughness (good understanding $\rightarrow$ efficient decision), and hesitancy (sufficient situation information

and good understanding → less hesitate to make a decision).

Drivers' action (steering, braking) can be used to assess whether they have good SA or not [29]. These objective measures assess specific task performance at decision points, which may be hard to interpret regarding SA. Common performance metrics include the quantity of output or productivity level, time to perform the task or respond to an event, and the accuracy of the response or, conversely, the number of errors committed [27].

The traditional model of human information processing proposed by [102] in their book *Engineering psychology and human performance* is shown in Figure 2.2. It provides a useful framework for analyzing the different psychological processes used in interacting with systems and for carrying out a task analysis.

Events in the environment are first processed by our senses and held briefly in short term sensory store (STSS) for no more than about a second. But sensation is not perception which involves determining the meaning of the sensory signal or event based on the past experience stored in our long term memory. In terms of visual sensory, sensation refers to *look* and can be captured by eye trackers but perception refers to *see*.

The dissertation mapped the SA measures to the different stages of the human information processing (HIP) framework, which is not a direct mapping to the traditional HIP by Wickens et al. [102]. Generally, the traditional HIP framework does not have provide sufficient level of detail associated with SA. Therefore the subsequent paragraphs discuss how this study maps the SA measure to the different stages of HIP framework.

Perception defined in the traditional HIP framework involves all three levels of SA defined by Mica Endsley which are perception, comprehension and projection. Some arguments for this conclusion are discussed as follows. First, they say "*perception* involves determining the meaning of the sensory signal" when defining *perception*, indicating they involve understanding the meaning of things in *perception*. Second, they say "perception and situation understanding do not always trigger an immediate response" which separates understanding as another process from *perception* but we can see there is no step between *perception* and *response selection* as depicted in Figure 2.2. This contradicts their definition of *perception* which includes the understanding of sensory signals. So there is the potential to find sub-steps in this *perception* step. Third, Mica Endsley distinguished comprehension and

projection by generating the understanding of the current status at the time t or generating the understanding of future status $t + \Delta t$. In other words, both comprehension and projection address the understanding of the sensory signal (involved in the definition of *perception*) but for different points in time. Therefore, we can conclude that Mica Endsley defined the sub-steps of this *perception* step by [102] as three SA levels – perception, comprehension, and projection.

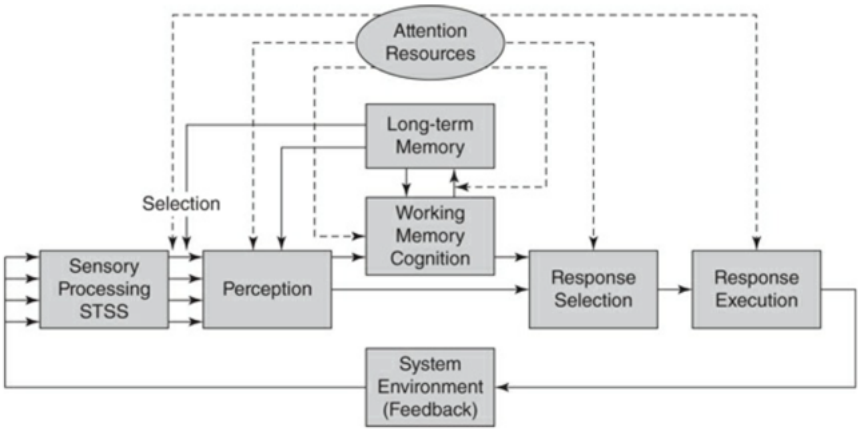


Figure 2.2: A model of human information processing stages [102]

2.2 Features Impacting SA

2.2.1 Environmental Features

Drivers situational awareness improves with repeated exposure to different environments. Environments with greater complexity (e.g., multimodal operations, multiple traffic signs, higher traffic density) can impair driving performance given increased workload [91, 24], which can then affect the operator's ability to take over in partially automated vehicles [63, 38].

There are varying complexity within a driving environment and they greatly depend on the environmental characteristics, object location, object type, and driver's eye movement [57, 109, 56]. Roadway features such as increased traffic density, visual clutter, and hazard occurrences have been shown to decrease driver's SA [40, 52]. Urban environments, with

greater environmental complexity, can also negatively impact driving performance when compared to rural environments [78]. The increased visual clutter can increase search time and error rates for driving-related information [41, 12]. Conversely, the lack of information or environmental cues may also result in drivers being less aware of their surroundings [55]. This complexity needs to be accounted for when estimating driver SA. Past studies have not provided as comprehensive an examination of different environmental properties. They may consider environments as two levels (easy, complex) [87] or rely more on visual differences between color, background, and scene density for their SA prediction model [109].

Road Geometries

Cues and hazards such as stop signs can easily be obscured by bushes or geometry. However, the driver could still glance at the Stop Sign, predicted perhaps by the roadway geometry to make the right decision [36].

Presence of Hazards

Drivers' hazard perception is defined as the ability to detect dangerous traffic situations [45]. The presence of hazards in the driving environment has been shown to cause degradations in high levels of driver SA [53]. However, when comparing hazard vs non-hazard conditions, hazards did not substantially affect the global awareness of the traffic situation [62]. Borowsky and colleagues showed that drivers rely on previous expectations of sign locations to determine where they will look for the signs. When these signs appeared in unexpected locations, experienced drivers made more errors in identifying these signs [8].

Road Users

A driving simulator study conducted by Jamson et al. [50] found that drivers in a fully automated vehicle allocated more attention to the road in situations with high traffic. High traffic density and therefore a greater number of objects lead to extended visual scanning and decision-making process overall resulting in less safe vehicle behavior (higher collision rate and shorter time to collision in takeover situations) [38, 85].

Visual Background Clutter

Environmental context and observer experience are generally critical in determining where observers allocate their attention [2]. Chapman and Underwood [13] demonstrated that visually less complex rural roads elicited the longest fixation durations and the shortest angular saccade distances while the opposite was found to be true for visually complex urban roadways. Young et al. [106] examined whether static roadside billboards disrupt driver situation awareness. Results showed that billboards did not appreciably affect the structure and content of situation awareness. Drivers typically directed attention to billboards only when driving demand was low.

Weather and Light Condition

The data is available for some weather-related conditions. However, it can be difficult to detect or perceive the visual information, such as poor visual information in darkness and poor weather, and poor visual conditions created by glare from runaway lighting [30].

Sound

The background sound of the environment helps the driver assess the distance or direction of objects in their surrounding. However, we have not found an approach that shows whether the driver has received the information from background sound that is similar to what is used for visual information (e.g., identifying whether the driver has attended to something based on their eye movement). Sounds are widely used to alter or send warnings as a supplement to the visual displays [5, 100].

2.2.2 Driver characteristics

Experience & Familiarity of the Environment

Driving experience and familiarity of the environment play important roles in levels 2 and 3 SA [34]. Endsley proposed that key cognitive attributes, attention focus, mental model, workload, and memory, are the main factors influencing SA. Each attribute has dominant impacts on different SA levels. A study in the aviation domain found that 76.3% out of

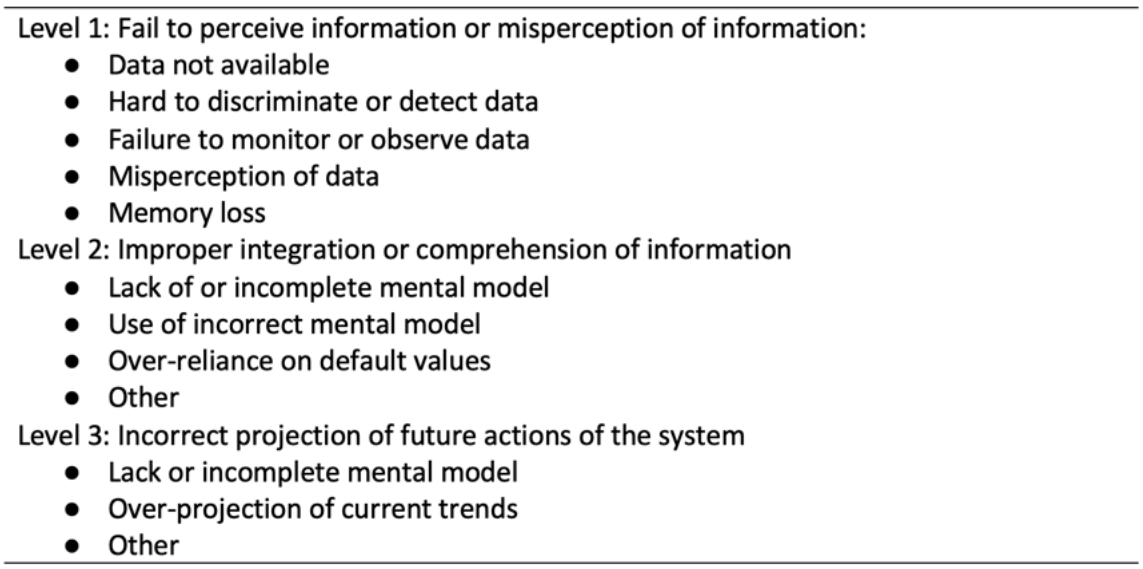


Figure 2.3: Taxonomy of Levels of Situation Awareness [51]

262 SA errors were classified as Level 1 SA errors, 20.3% were Level 2, and 3.4% were Level 3 [51]. The detailed causations are listed in Table 2.3. Attention focus and memory mainly impact level 1 and mental models mainly affect levels 2-3.

Rouse and Morris [89] defined mental models as "mechanisms whereby humans are able to generate descriptions of system purpose and form, explanations of system functioning and observed system states, and predictions of future states." The mental models typically contain information about not only the components of a particular system but also how those components interact to produce various system states and events. Thus, mental models can provide much of the higher levels of SA (comprehension and projection) without loading working memory [34]. Mental models highly correlate with experience as experience allows the development of mental models and schema, pattern matching between the perceived elements in the environment and existing schema/mental models can occur on the basis of pertinent cues that have been learned [34].

Age

Due to age-related declines in sensory functions, older drivers were expected to exhibit significantly lower level 1 SA (perception) than young drivers [1]. Related to this, older drivers were also expected to exhibit lower level 3 SA (projection) than young drivers [53].

2.3 Predicting Driver SA

Few studies attempted to monitor and predict SA in real-time. The current research relies on direct SA measures, like SAGAT [28] and situation awareness rating technique (SART) [97], which pose operational difficulties in real-life applications. SAGAT requires participants to answer questions at three levels, which can interrupt the task's continuity, while SART disrupts the ongoing task of the driver by asking questions during or after the task.

In recent studies, the use of process measures, particularly eye-tracking data, is being explored for the dynamic assessment of SA, Eye-tracking data has been shown to be objective and capable of inferring SA, making it an ideal alternative to direct SA measures, according to previous studies [56, 57]. To predict driver SA from eye-tracking data, these studies require multiple SA measures to be recorded in the same experimental settings, often necessitating the development of new studies to obtain eye-tracking data and ground truth measurements of driver SA using direct measures.

There have been four studies in the driving domain related to predictive models of SA arranged in chronological order in Table 2.1, which involve Honda Research Institute (HRI) US working in collaboration with Virginia Tech Transportation Institute (VTTI) [56, 57, 109] and University of Washington (UW) [104, 80]. The remaining two studies were conducted one at the University of Michigan (UMich) & Delft University of Technology (TU Delft) in Netherlands [62, 108], and the other at Technical University of Munich (TUM) & BMW both in Germany [42, 43].

HRI & VTTI's first paper [56] proposes a new method of using real-world driving videos in measuring drivers' SA and investigated the relationship between participants' visual attention and awareness of road hazards. Road hazards in this study refer to objects (pedestrians and other vehicles) that are risky to the driver. The authors propose a video-mimicking

Table 2.1: Studies on Predicting Driver Situation Awareness

Team	Study	Year(s)	Driving Environment	Scenarios	Ps.∗	Obs.†
VTTI & HRI	[56]	2019	CG-based, CG-mimicking, and Video-mimicking	Intersections	21	486
	[57, 109]	2020, 2021	Video-mimicking	Intersections	44	1232
UMich & TU Delft	[62, 108]	2020, 2021	CG-mimicking	Freeway	32	896
TUM & BMW	[42, 43]	2020, 2021	Video-mimicking	8 specific	8	64
UW & HRI	[104, 80]	2022	Video-mimicking	75 unrestricted	40	6240

∗ Ps. - the count of participants used for data analysis with outliers and incomplete instances excluded.
† Obs. - the abbreviation for observations.

method where participants mimic driving while watching pre-recorded videos from real-world driving. Drives were paused, an image as shown in Figure 2.4-bottom with road hazards removed using Adobe Photoshop from the original scene shown in Figure 2.4-top and zones of interest overlaid was displayed on the simulator screen. An experimenter then asked drivers in which zones they detected road hazards and their confidence in those answers. Drivers' SA score was assigned as the confidence for only correctly recognized road hazards. Eye-tracking data was collected and synchronized with the videos, i.e., coordinates of drivers' gaze were mapped to the videos. They also collected and compared the data using computer graphic (CG)-based method (drive the same scenarios in a driving simulator) and CG-mimicking method (mimic driving the same scenario in a driving simulator). The results of this work supported the validity of the proposed method: 1)no differences between CG-based and CG-mimicking were detected suggesting mimicking the task does not impact drivers' gaze behavior 2) comparison between CG-mimicking and video-mimicking indicates that the latter is as comparable for driver visual detection. The study also found with a multilevel regression model that the gaze measures explain only 44% variability of driver SA, indicating the need for involving other predictors for SA such as properties of objects and the environment.

In HRI & VTTI's second paper [57], they doubled the sample size from 21 to 44 and focused only on experiments with video-mimicking simulation environments. Two SA measures were analyzed: 1) binary SA score equals whether correct SA is generated and 2) SA confidence ranging from 0 to 1. Correlation and regression analyses for SA measures

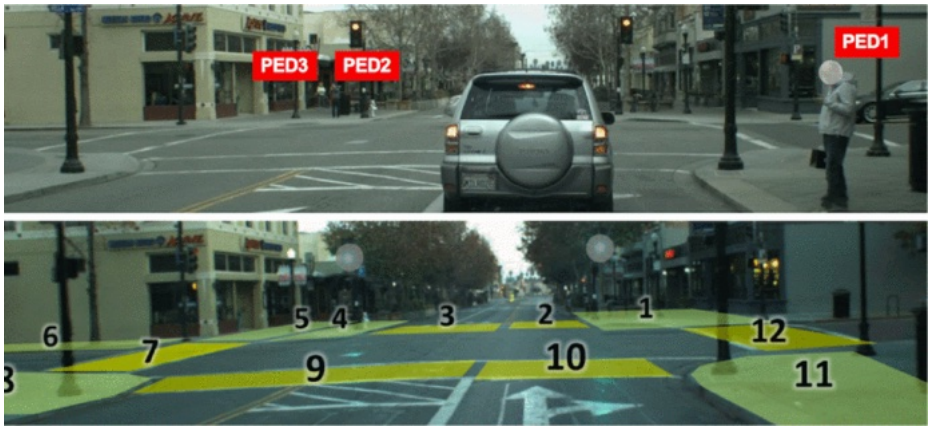


Figure 2.4: Example of the technique for querying SAGAT of pedestrians with top shows an actual driving situation and bottom shows the zones of interest presented at a pause [56]

Figure 2.5: The driving simulator used in study [56] with a projection screen that can switch between simulation environments including CG-driving, CG-mimicking and Video-mimicking

and eye-tracking data were conducted from three levels: 1) event-wise analyses were conducted on data collected from each driver for each road hazard; 2) object-wise analyses were conducted on data averaged across drivers for each road hazard; 3) driver-wise analyses were conducted on data averaged across road hazards for each driver. Object-wise analyses showed the best results and supported their conclusion in the first paper [56] that eye-tracking data can be a promising insufficient predictor of driver SA.

In the third paper [109], HRI focuses on building a predictive model for object-wise binary driver SA score and largely expanded the predictors, including driver gaze features, object properties, and characteristics of human visual sensory and memory mechanism. The authors compared Support Vector Machine (SVM) models with different sets of inputs including eye gaze features, object properties, human visual sensory-based features, and

memory features achieving the best prediction accuracy of 72.4% compared to a chance rate of 55.8%. This study was well designed with a clear structure for selecting the features that are used for the predictive model. They validated the method of using video-mimicking for SA measure collection and showcased a promising approach. More specifically, using eye-tracking features with other features such as object properties to predict driver SA. However, there were three major limitations. The first is about the SAGAT query design. It's hard to tell which road hazard the participant detected when there are multiple hazards in one zone of interest. Second, all scenarios are at intersections so there is a need to test in more general road scenarios. Third, most drivers are very young with an average age of 26.5 and an SD of 7.2. Previous studies showed that older drivers are expected to exhibit significantly lower level 1 and 3 SA (perception and projection) than younger drivers due to age-related declines in sensory functions but higher level 2 (comprehension) due to richer experience [1, 53]. Thus, there is a need to consider various age groups in future studies. Our work tries to break those limits and will be described in detail as preliminary research in Chapters 4 and 5.

The study by UMich & TU Delft [108] measured driver SA during takeover transitions using a CG-mimicking simulation environment as shown in Figure 2.6. The data used are from an existing study [62] by the TU Delft, which describes experimental design and data collection work in detail. The duration of the scenarios varies from 1 to 20s which is very short and only freeway scenes were tested. For each scenario, 5 to 6 vehicles were added as surrounding road users. Data were collected from 33 participants between 22 to 29 (mean = 24.2, SD = 1.8) including 29 males and 3 females. Participants place surrounding vehicles via an interface shown in Figure 2.7. They first choose from "faster" "equal" or "slower" vehicle symbols indicating the relative speed to their ego car and place the car in the lane at the distance they perceived. The measurement of scenario-wise SA was computed by normalizing three error scores inversely into a scale from 0 to 1 and summing with equal weights. Scenario-wise is defined against the event-wise, object-wise, and driver-wise described in HRI & VTTI's study, meaning to average data across objects in each scenario for each driver. The three error scores include 1) the absolute difference between the true number and participant-placed number of vehicles, 2) the error percentage of the distance

between correctly placed vehicles and the true vehicles, and 3) the total speed difference between correctly placed vehicles and the true vehicles. Performance in RMSE, MAE, and correlation coefficient between the predicted SA and the ground truth were compared across selected machine learning models with the same predictors. Light Gradient Boosting Machine (LightGBM) performs best for all three performance measures: RMSE range of different models 0.121-0.108; MAE range 0.097-0.085; and Correlation coefficient range 0.436-0.801.

$$\begin{aligned} SA = & \frac{1}{3}|N_{true} - N_{placed}| + \\ & \frac{1}{3}\frac{|dist(true,\ ego) - dist(placed,\ ego)|}{dist(true,\ ego)} + \frac{1}{3}\sum_{veh\ i\ in\ s\ j}|V_{true,\ i} - V_{placed,\ i}| \end{aligned} \tag{2.1}$$

Figure 2.6: CG-mimicking simulation environment used for study [108, 62]

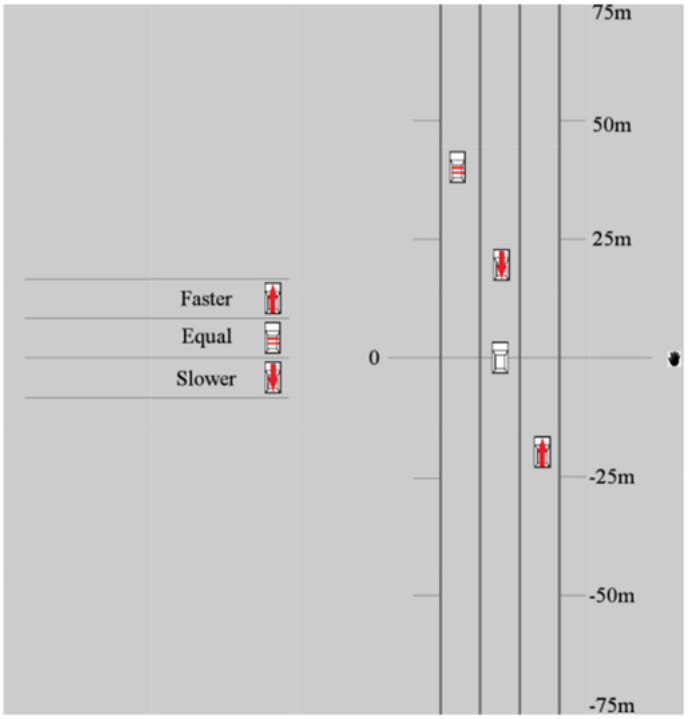


Figure 2.7: SA response recorder interface [108, 62]

UMich & TU Delft's study [108] explores a different granularity of SA, scenario-wise compared to HRI & VTTI's work. This is because of their different research purpose. HRI's study aims to selectively provide assistance using ADAS when the system detects road hazards drivers are not aware of. UMich & TU Delft [108]'s paper aims to generate drivers' SA for the entire scene in real-time and warn drivers when the SA score is low. That is, the study will only warn the drivers to get their attention back but the HRI's study moves one step further to also provide hints for where to direct their attention. Small sample size, narrow age range, and unbalanced gender group are limitations in the data. The SA measuring method requires mental rotation: participants experienced a forward-perspective driving situation but had to position the cars from a top-down perspective, which may increase the difficulties in answering the questions making the results reflect driver SA plus their ability to do mental rotation.

The study of TUM & BMW [43] leans more toward the application of SA predicting

technology and aims to develop a system that can predict driver SA in real-time. Inspired by a previous study by Hooey et al. [44] in the context of aviation, this study defined SA score at time t as the ratio of actual SA and optimal SA:

$$SA_{ratio}(t) = \frac{SA_{actual}(t)}{SA_{optimal}(t)} \tag{2.2}$$

The optimal SA defines the SA when the operator perceives and comprehends all situation elements, which refer to pieces of information that are necessary to support the operator's high-level tasks [93]. In the context of aviation, SEs are mainly static cockpit instruments visualizing the state information of the plane. However, in the transportation field, SEs are mainly dynamic and outside the vehicle, which is similarly defined as the road hazards in HRI & VTTI's study [56, 57, 109]. The way Hooey et al. [44] measure optimal SA for pilots as the weighted sum of m required and n desired SEs (see Equation 2.3) is not applicable for drivers. Thus, TUM & BMW's study used predicted region of interest (ROI) as the optimal SA with the model created in their previous study [42].

$$SA_{optimal} = \sum_{r=1}^{m} 2 \cdot SE_{opt,r} + \sum_{d=1}^{n} SE_{opt,d} \tag{2.3}$$

The actual SA defines the current awareness status of the operator. TUM & BMW didn't use questionnaire-based SA measurements as in the other three studies. Data needed was collected from 8 participants driving in 8 unique 10-second real-world scenarios displayed on a computer screen. Eye gaze data was captured using an eye tracker. For each SE, perception level was categorized as "detected" if one gaze data point locates in its ROI, "comprehended" if more than one gaze point locates in its ROI, and "undetected", otherwise. Actual SA was then calculated as a weighted sum of required and desired SEs multiplied by the perception level:

$$SA_{actual} = \sum_{r=1}^{m} 2 \cdot p_{irt} + \sum_{d=1}^{n} p_{drt} - o \cdot \gamma \tag{2.4}$$

Perception level p_{drt} and p_{irt} both are assigned as 0, 0.5 and 1 for "undetected", "detected" and "comprehended" SE respectively. This Equation 2.4 improved Hooey et al.'s [44]

by adding the term $o \cdot \gamma$ to punish the non-ROI gaze clusters.

Finally, ROI prediction was concluded as a proper definition of the optimal SA and the way of mapping eye gaze points to ROIs as a proper measure for actual SA based on descriptive analyses. The study tried to consider higher-level SA (comprehension) and considered the distraction of non-ROIs, which are very innovative. However, there are a couple of limitations. First, both the number of scenarios and participants are very small. Second, the weight for each actual SA category was inherited directly from the study in the aviation field, which may not be applicable to drivers. Third, the study made the conclusion based on descriptive analyses without using more rigorous statistical methods. Forth, although the SA for each SE can be computed, the authors computed SA as scenario-wise scores, which will not provide selective assistance for drivers. This last limitation is shared by the study of UMich & TU Delft.

In all, HRI's study is more comprehensive and structural compared to the rest. The scenarios are longer which enables the participants to feel more immersed. HRI's study also has the richest features and most data points so far. Common limitations shared by the studies include:

1. Modeled only level 1 SA – perception. However, learning higher levels of SA is essential and beneficial. For example, perceiving a road user correctly but failing to predict its trajectory will still raise safety concerns.

2. Modeled scenario-wise SA instead of object-wise. In this case, they will not be able to selectively provide assistance to warn drivers only of the hazards they should be they are not aware of.

3. All studies relied on eye-tracking data to predict SA. This will be hard to apply since eye trackers are expensive, need calibration and are always used in lab environments, and may cause some privacy concerns.

4. The driving scenarios of previous studies are very limited and there is a great need to build a predictive model of SA that works for various environments.

This study aims to inherit the strengths of previous studies and address the research gaps listed above.

Chapter 3

RESEARCH AIMS

The primary goal of the study is to provide assistance to drivers by only alerting them of hazards they may not be aware of. This goal includes two parts: 1) estimation of object importance in the traffic scene and 2) estimation of driver situation awareness. This study focuses on the second part and the object of interest which is the output from the first part was manually labeled and evaluated by research assistants. This work will address four aims listed below and Figure 3.1 shows how the components are connected in this study.

Study Aim 1. Develop a method for effectively capturing driver SA.

Previous studies have employed query-based SA assessments, which capture SA as a binary outcome based on a comparison with the actual situation [35]. This approach involved inquiring about the participant's observation of specific elements in the scene. Kim et al. [56, 57] took this approach a step further by asking SAGAT queries at intersections about which areas participants had observed road users, and evaluating their SA scores based on the presence of potential hazards at those intersections. However, this method had limitations in cases where multiple potential hazards were present in the same area, making it challenging to determine which hazard the participant was aware of. This study advances the method by capturing the exact location in the scene where participants have observed road users.

Study Aim 2. Identify potential predictors of driver SA.

Since the study adopts the driver SA capture approach outlined in Aim 1, it is essential to determine how to assess SA using the data gathered from the experiment as well as understanding the influence of environmental, object, and driver features on this self-defined

SA measure before proceeding to train the predictive model for driver SA. Additionally, this process aids in feature selection for subsequent model training in Aims 3 and 4.

Study Aim 3. Predict drivers' SA using eye-tracking data

Eye-tracking data captures where the eyes "look," but not necessarily what the driver truly "sees," which is indicative of SA. Therefore, having eye-tracking data does not necessarily equate to understanding the driver's perception accurately, highlighting the importance of a predictive model for SA based on eye-tracking data. The features analyzed in Aim 2 can complement the features extracted from eye-tracking data in predicting SA.

Study Aim 4. Predict driver SA using real-world driving videos.

Previous studies have relied on eye gaze data collected by eye trackers as input for predictive models of SA. However, wearing eye trackers can be uncomfortable, especially for individuals with vision impairments, and they are also costly, mainly used in controlled laboratory environments. Although facial videos can serve as an alternative to gather eye-gaze information, privacy concerns arise. Given that drivers heavily depend on visual cues and ADAS systems often incorporate cameras capturing surrounding videos, predicting SA directly from this video data would be advantageous. Hence, this study suggests utilizing visual attention predictive models to estimate eye fixation and subsequently predicting SA based on these estimated gaze-related features. The features analyzed in Aim 2 can complement the estimated gaze-related features in predicting SA as well. Model results will be compared with those from Aim 3 to evaluate the possibility of using the estimated gaze-related features as an alternative to the gaze-related features extracted from the eye-tracking data.

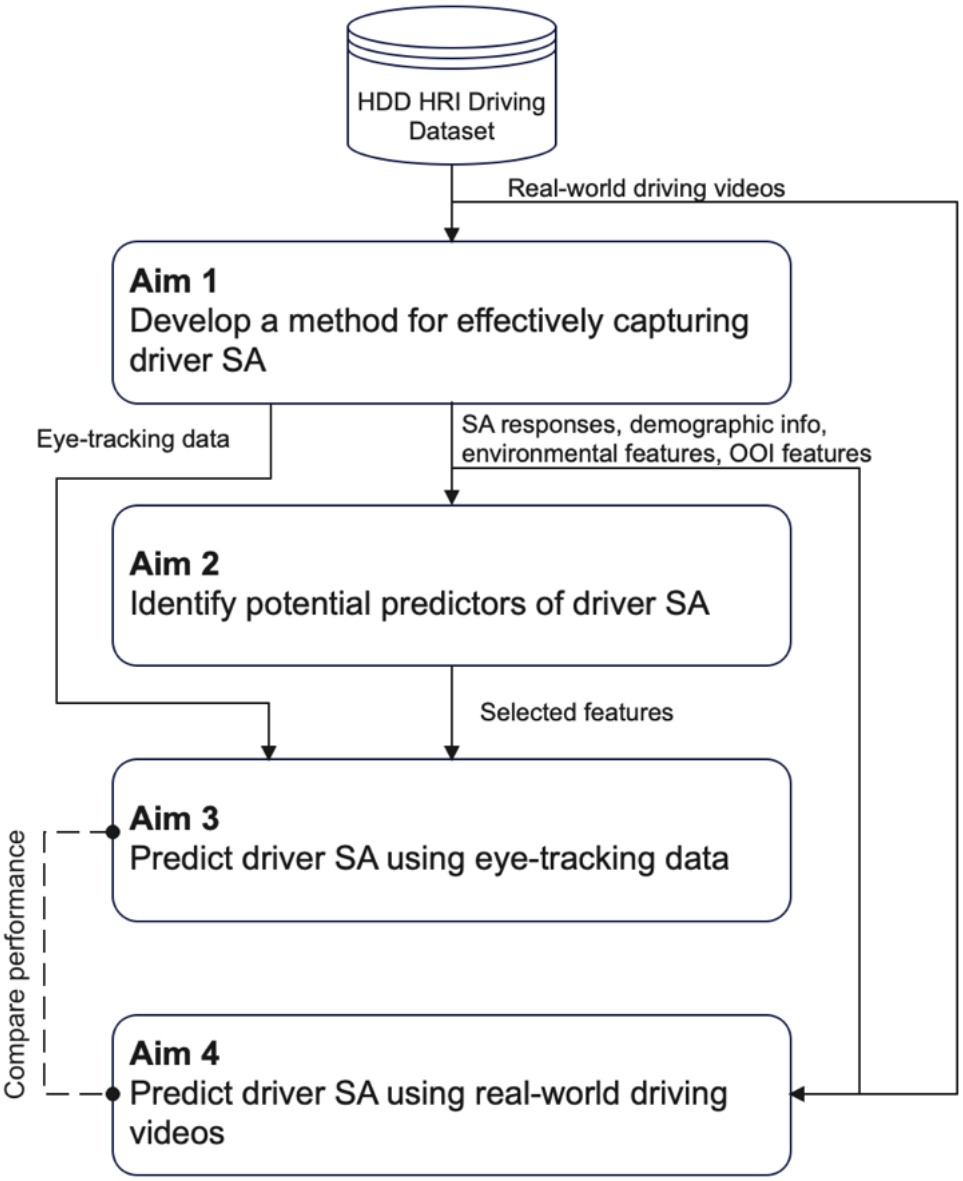


Figure 3.1: Study aims and their relationship

Chapter 4

AIM 1: NEW METHOD FOR CAPTURING DRIVER SA

Aim 1 proposes a new method for collecting driver SA. Data is used to capture direct SA measure (SAGAT response) and indirect SA measure (eye tracking data) collected in one controlled study.

The findings of this capture are already published in several papers [104, 80]. The study has also been approved by the University of Washington Institutional Review Board (IRB) Committee (No: STUDY00013730).

4.1 Participants

A total of 40 drivers were recruited for this study based on Table 4.1. Younger drivers were screened for the age range 18-40 and older drivers included those older than 40 years old. Pairwise t-tests showed no age differences within the same gender group. Younger drivers had 6.6 years of driving experience on average (SD = 4.8 years) and older drivers had 32.3 years on average (SD = 8.0 years).

Table 4.1: Participant demographics

	Total	Younger & She	Younger & He	Older & She	Older & He
Data collected	67	13	16	18	19
Data used	40	10	10	10	10
Age range	18-61	18-40	20-33	41-60	41-61
mean	36.7	22.7	24.0	49.3	50.9
SD	14.88	6.46	3.65	6.57	7.95

Participants were recruited using the University of Washington email lists and from online sources such as Facebook and Craigslist posts around Seattle, WA. 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. Eleven drivers had normal eyesight and

nine used glasses or contact lenses daily (diopter within the range of -5.0D to +3.0D). Data results of 27 out of 67 drivers had to be excluded due to motion sickness or had unusable data records due to technology issues, see Table 4.1 for more detail. Upon successful completion, the drivers were compensated US$37.50 for their time in the study.

4.2 Apparatus

4.2.1 Simulator Environment

This study used the hardware of a fixed-based driving simulator with real-world driving scenes shown across three 42-inch plasma displays (1,280 × 720 resolution) as shown in Figure 4.1 The setup was designed to be comparable with other immersed driving simulation environments in terms of manual driving demand, participants' gaze behavior, and awareness of road hazards [56]. 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.

Figure 4.1: The driving simulator of HFSM lab showing a scenario across three displays

4.2.2 Driving Scenarios

The videos used were from HDD HRI Driving Dataset[1], recorded in urban environments in daylight conditions in San Francisco, California, with a human driver using an instrumented vehicle [86]. There were no rural or highway roads recorded. The videos presented various densities of moving road users such as vehicles, pedestrians, and cyclists.

The visual complexity of the 2D scenario images was considered when selecting scenarios for our experiment, which is calculated as the ratio between the compressed and uncompressed image file size following the equation below.

$$\text{Visual complexity} = \frac{\text{Compressed size}}{\text{Original size}} \tag{4.1}$$

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) [67]. Theoretically, this value ranges from zero to one. Greater values indicate higher image visual complexity. In our study, the range for the video images is from 0.2 to 0.7 (Mean = 0.43, SD = 0.10).

This measure allows us to capture the visual complexity of the environment. For example, visual clutter or contrast in the scene caused by shadows or buildings may impact the road user's ability to accurately perceive pedestrian behavior. Scenarios with greater visual complexity tend to have more road users and buildings, and there is greater contrast between objects.

Event videos were classified into five levels by visual complexity as [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 Figure 4.2. Video clips were manually selected from each complexity level according to the proportion: 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, accounting for 0.55% of all, was filtered out from the eligible video pool).

[1]Honda Research Institute Driving Dataset: https://usa.honda-ri.com/hdd

Figure 4.2: Examples of scenarios with different image visual complexity

4.2.3 Touch Screen Recording

A 13-inch iPad was used to record participants' touch responses (Figure 4.3). The iPad display would first show the same driving scene as the scenario displayed in the simulator. However, some objects were purposely removed. This included 1) pedestrians, vehicles, and cyclists on the road; 2) vehicles that may potentially move; 3) pedestrians close to the road; 4) pedestrians, cyclists, and vehicles that are on the same block as drivers.

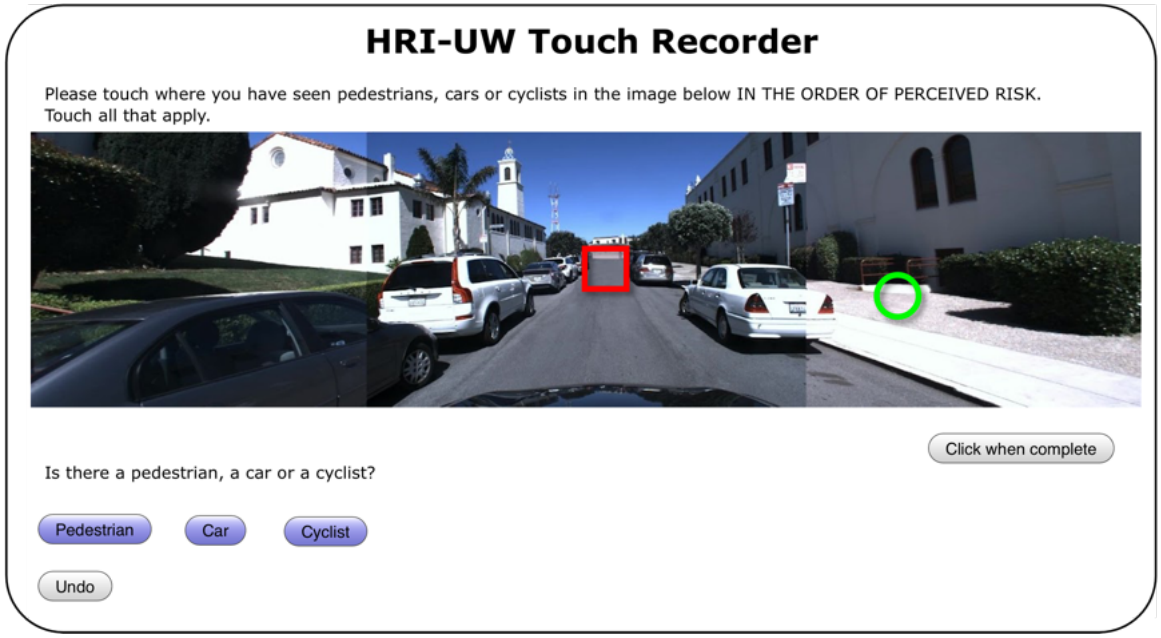
These objects that were removed were identified as objects of interest (OOI). Non-OOI (e.g., vehicles parked on the side of the road) were considered less risky compared to OOIs and were not removed from the touch recorder image.

When the scenario displayed on the simulator was paused, drivers were instructed to pick up the iPad and mark the locations and types of every detected OOI observed on the scene using either a stylus or their finger. 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 changes 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.

(a) Scene

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



(c) Touch recorder - marked scene

Figure 4.3: Examples of touch recorder display

4.2.4 Eye Tracker

The Tobii Pro Glasses 3 eye tracker[2] was calibrated for our study to track participants' eye movements while driving. Participants were highly recommended to wear contact lenses if they have any type of the following vision problems:

- Near-sighted: can only see items that are near without glasses;
- Far-sighted: can only see items that are far away without glasses;
- Near-and-Far-sighted: cannot see items that are near or far without glasses.

For participants who have the vision problems above but do not have contact lenses, we tested with corrective lenses added to the eye tracker with a diopter from -0.5 to +3.0.

Figure 4.4: Tobii Pro Glasses 3 eye tracker

4.3 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 non-female, older female, and older non-female). Each of the whole plots included 10 participants, i.e., 10 subplots as shown in Table 4.2. P1 to P10 refers to the ten participants in each subplot. Each whole plot was run as a balanced incomplete block design. In this study, a total of 75 scenarios were used which were split

[2]https://www.tobiipro.com/product-listing/tobii-pro-glasses-3/

into five sessions with 15 scenarios per session. Each cell in Table 4.2 refers to the index of a scenario from 1 to 5. Each participant experienced two out of five sessions or 30 scenarios. Each session was replicated eight times, accumulated for the whole experiment, by different participants in the same whole plot.

Table 4.2: Session combinations experienced by each participant (subplot) in each whole plot.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
1st Session	3	4	2	2	1	1	4	5	3	5
2nd Session	5	1	4	3	3	2	5	1	4	2

4.4 Data Collection Procedures

The experiment includes four main procedures: (1) informed consent and a pre-experiment survey, (2) training on the procedures and apparatus for the experiment, (3) one practice drive, and (4) two study drive sessions.

Participants were required to read through a consent form in hard copy and sign if agreed to run the study. Upon informed consent, a pre-experiment survey was provided, querying the demographic information and driving experience of the participant. Participants were informed that they can skip any question that they feel uncomfortable answering.

Upon completion of the survey, an experimenter will introduce the procedures of the study with slides and show participants how to interact with all the apparatus used in the study including the touch recorder, the driving simulator, and the eye tracker. After that, the experimenter will guide the participants to sit in the driving simulator, put on the eye tracker, and test the calibration of the eye tracker. Corrective lenses were applied to the eye tracker for short or far-sighted participants who came with frame lenses.

Before the two study drive sessions, participants were trained with a 5-minute practice drive containing two scenarios to get familiar with the driving simulator operations. Though participants' operations will not give feedback to the simulator, they were highly recommended to use the pedals, turn signals, and the steering wheel as they usually drive to match the movement of the vehicle in the displayed video, in order to get more immersed

in the scenario. The video was paused randomly during each scenario for SAGAT queries so that the drivers would not predict when the video would pause to get prepared. At each pause, participants picked up the iPad and answered SAGAT questions using the touch recorder web App. Between the two study drives, participants took a 10-minute break.

Each study drive session took about 20 minutes and the entire experiment took about 75 to 90 minutes. Considering possible motion sickness issues, we informed participants that they can pause or stop the study at any time if they feel sick or uncomfortable.

4.5 Data Collected

The data collected from the experiment include drivers' demographic information from the pre-experiment survey; SA responses recorded as the coordinates and types of objects for each scenario from the web-based touch recorder; and drivers' eye gaze data from the eye tracker.

Chapter 5

AIM 2: PROMISING PREDICTORS OF DRIVER SA

The content of this chapter has already been presented in conference papers [104, 80] and published in journal article [103].

Features excluding the eye gaze features will be discussed in this chapter from two perspectives: object-wise and scenario-wise. Though using different SA score computing methods, analyses from both perspectives presented in previous studies have shown strong relations between object-wise or scenario-wise SA with gaze features. Hence, SA for both perspectives was analyzed in this study. Moreover, since the final goal of this study is to build a model to predict SA in order to provide selective assistance using ADAS, object-wise analyses will be our main focus.

This chapter will focus on features other than gaze features. There are three reasons for this arrangement: 1) this study has relatively larger scenario and OOI sets and would like to explore more possible predictors from environmental and object properties; 2) eye gaze features will undoubtedly be used for building predictive models using eye tracking data, thus will be more appropriate to discuss for the data preparation for those models; 3) the eye gaze data have been mapped to the driving video so far but feature extraction procedures have only been tested on a small set of data, so it would be more clear to discuss this content in the section for proposed work.

5.1 Object-Wise SA

5.1.1 Dependent Variable

SAGAT Responses is a categorical variable with 4 levels: hit, trajectory-delayed perception, trajectory-perception, and miss.

The content below describes how it was computed in detail.

The frame at pause is called the pause frame with index 0 assigned. Critical objects in

seven frames including the pause frame and six frames with indices ± 30,± 20, ± 10 were labeled for each scenario using tool *labelme*[1]. The seven bounding boxes are represented by the boxes with solid borderlines in Figure 5.1. Polynomial interpolation with a degree of three was applied to interpolate the object's trajectory based on the manually labeled bounding boxes. Specifically, the trajectory of the object's center point was interpolated together with the width and height of the bounding boxes. These interpolated bounding boxes along the object's trajectory are depicted as boxes with dashed borderlines in Figure 5.1. Figure 5.2 shows an example of overlaying the bounding boxes to a scenario's pause frame. In Figure 5.2(a), only bounding boxes manually labeled for the pause frame are overlaid. All seven manually labeled bounding boxes are depicted in Figure 5.2(b). The estimated trajectories of the bounding boxes overlaid to the pause frame are shown in Figure 5.2(c). The colors red and green indicate pedestrians and vehicles respectively.

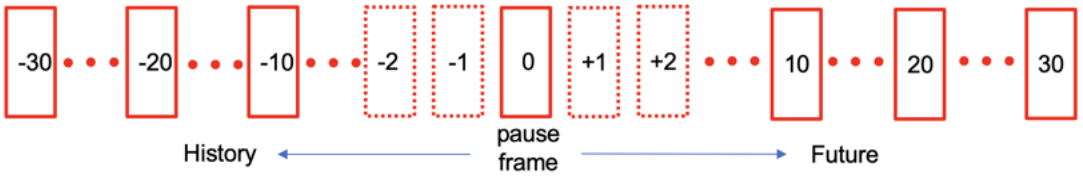


Figure 5.1: Indices of manually labeled and interpolated bounding boxes

We defined the results of the participant's SAGAT response for each object as one of four categories:

- *Hit.* Touch located in the bounding box of the pause frame was recognized with a matched response for the type of object. The pause frame is the frame at pause for the SAGAT query depicted as box 0 in Figure 5.1.
- *Trajectory-Projection* and *Trajectory-Delayed Perception.* Both categories define touches out of the bounding boxes in the pause frame but were in one of the labeling boxes along the estimated trajectory. If the touch locates in the bounding boxes $[-30, 0)$ in

[1] https://github.com/wkentaro/labelme

Figure 5.1, it will be categorized as Trajectory-Delayed Perception, and if the touch locates in the bounding boxes $(0, 30]$, it will be categorized as Trajectory-Projection. The index of the frame closest to the pause frame measures the amount of delay in perception or the amount of advance in projection.

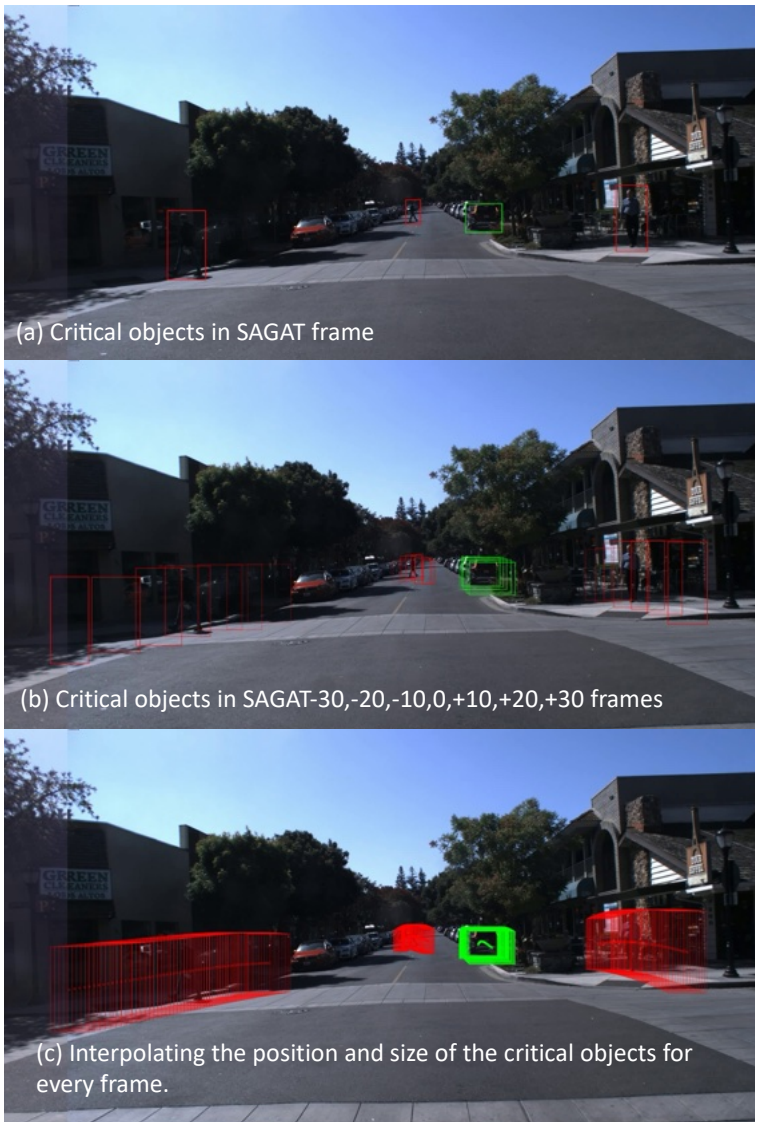


Figure 5.2: Example for overlaying manually labeled bounding boxes and objects' estimated trajectories to the pause frame

- *Miss.* No touch locates in any bounding boxes, including the bounding box in the

pause frame and the bounding boxes along the estimated trajectory.

5.1.2 Independent Variable

- *Object Type.* There were 3 levels for object type: pedestrian, car, and cyclist.
- *Size of Object.* Area of the labeling box for each object in pixels.
- *Visual Image Complexity.* Visual image complexity of the pause frame was calculated using Equation (4.1).
- *Number of Objects.* The number of objects in the pause frame for each object type. The objects were manually counted and then verified twice by two research assistants.
- *Gender.* Gender of participants, 2 levels, male or female.
- *Roadway Type.* Roadway type included three levels: segment, intersection, and roundabouts.

5.1.3 Data Analysis Method

A multinomial logistic regression model was used to predict the likelihood that the participant provided a hit, trajectory-projection, trajectory-delayed perception, or miss when generating the situation awareness in various driving scenarios. Given that there are four levels, the model predicts the odds of being in one category when compared to the other three; this results in six unique comparisons. Three of the comparisons are shown in Table 5.1 (1) hit compared to miss, (2) trajectory projection compared to miss and (3) trajectory-delayed perception compared to miss.

5.1.4 Results

There were 390 critical objects in the 75 scenarios, each experienced by eight young drivers. The objects consisted 137 cars, 240 pedestrians and 13 cyclist. For each of the $390 \times 8 = 3,120$ observations that whether the participant perceived the object, perceived the trajectory of the object or miss the object, there are 152 Hits, 190 Trajectories and 2,778

Miss. In terms of the 390 objects, 101 objects were correctly perceived (Hit) and 117 objects' trajectories were perceived (Trajectory).

The multinomial logit model shows that all but object size was significant at $\alpha = 0.01$. Object size was not significantly different between Trajectory Projection vs Miss (Comparison II). In terms of object type, cyclists were less likely to perceive (hit) and less likely for drivers to perceive their trajectories compared to cars and pedestrians. Pedestrians were less likely to be correctly perceived (OR = 0.539) but drivers seemed more likely to perceive their trajectories compared to cars (OR = 1.03, 1.07). Smaller objects were more likely to be correctly perceived (OR = 3.89) and it seems drivers were more likely to project the smaller objects (OR = 1.26) rather than having delayed perception (OR = 0.39). Females were more likely to correctly perceive the objects (Or = 1.17) and the trajectories (OR = 1.40, 1.14) compared to the non-females. Increased visual complexity seemed to make it more likely for the drivers to perceive the object (OR = 1.00) but less likely to perceive the trajectories (OR = 0.44, 0.31). The increased number of objects in the scenario makes it more likely to miss an object through all the comparisons (OR = 0.98, 0.97, 0.98). Compared to intersections, drivers were less likely to perceive objects on roadway segments (OR = 0.879), but more likely to perceive in the way of their trajectories (OR = 1.16, 1.37). Objects and their trajectories were more likely to be perceived at roundabouts (OR = 2.10, 1.21, 1.57).

5.1.5 Findings

Having touches recognized as "trajectory" and "hit" suggests that participants sometimes perceive trajectory instead of the object's exact location. Cases of detecting trajectories before and after the pause frame were both detected with the number of frames and time ahead or delayed, indicating that this proposed SA collecting method has the potential to provide quantitative measures of higher-level SA – comprehension and prediction.

The environmental properties such as number of objects, object size, visual complexity, roadway type were tested to have significant impact. In terms of individual property, females are more likely to correctly perceive objects and trajectories compared to the non-females.

The results showed that it is reasonable for us to incorporate this information into predictive models for driver situation awareness in order to warn the drivers of the objects that they are not aware of but they should be aware of (avoid redundant warnings).

Table 5.1: Findings of multinomial logit model predicting the likelihood of SA responses

Comparison	I: Hit vs. Miss				II: Trajectory-Projection vs. Miss				III: Trajectory-Delayed Perception vs. Miss			
Variable	Coeff.	S.E.	OR**	95% CI	Coeff.	S.E.	OR	95% CI	Coeff.	S.E.	OR	95% CI
Intercept	-2.71*	0.46	0.07	(0.03, 0.17)	-2.71*	0.55	0.07	(0.02,0.20)	-2.81*	0.57	0.06	(0.02,0.19)
Object type:												
car (ref, 35%)												
cyclist (3%)	-13.91*	0.00	0.00	(0.00, 0.00)	-19.90*	0.00	0.00	(0.00, 0.00)	-15.01*	0.00	0.00	(0.00, 0.00)
pedestrian (62%)	-0.62*	0.18	0.54	(0.38, 0.76)	0.03*	0.22	1.03	(0.67, 1.58)	0.07*	0.23	1.07	(0.68,1.67)
Object size/10^6	1.36*	0.31	3.90	(2.11, 7.19)	0.23	0.60	1.26	(0.39, 4.06)	-0.94*	0.30	0.39	(0.22,0.70)
Gender:												
Male (ref, 50%)												
Female (50%)	0.16*	0.17	1.17	(0.84, 1.63)	0.34*	0.21	1.40	(0.93,2.11)	0.13*	0.21	1.14	(0.75,1.73)
Visual complexity	1.00*	0.81	2.72	(0.56, 13.32)	-0.81*	0.98	0.44	(0.07,3.01)	-1.19*	1.00	0.31	(0.04,2.18)
# of objects	-0.02*	0.01	0.98	(0.96, 1.01)	-0.03*	0.01	0.97	(0.94,1.00)	-0.02*	0.01	0.98	(0.95,1.01)
Roadway:												
intersection (ref, 31%)												
segment (67%)	-0.13*	0.19	0.88	(0.61,1.27)	0.14*	0.25	1.16	(0.73,1.84)	0.31*	0.24	1.37	(0.84,2.24)
roundabout (2%)	0.74*	0.46	2.10	(0.85,5.19)	0.19*	0.65	1.21	(0.34,4.33)	0.45*	0.66	1.57	(0.43,5.71)

Number of Obs.: 3120 AIC: 2854.16 Log Likelihood: -1400.08 (df = 27) * $p < .001$ ** OR: odds ratio

5.2 Scenario-Wise SA

Dependent Variable

An SA score, represented as a percentage, was computed using the proportion of objects of interest (OOIs) correctly identified in each scenario:

$$\text{SA score (\%)} = \frac{N_{OOI} \text{ answered correctly}}{\text{Total } N_{OOI}} \times 100\% \tag{5.1}$$

An answer was marked correct if the participant's touch position was located within the bounding box (see Figure 5.3). A touch position located within the bounding box but marked as the wrong type of object or touch outside the above area was considered a wrong answer. Wrong answers or missing the object (not answering) did not penalize the SA score.

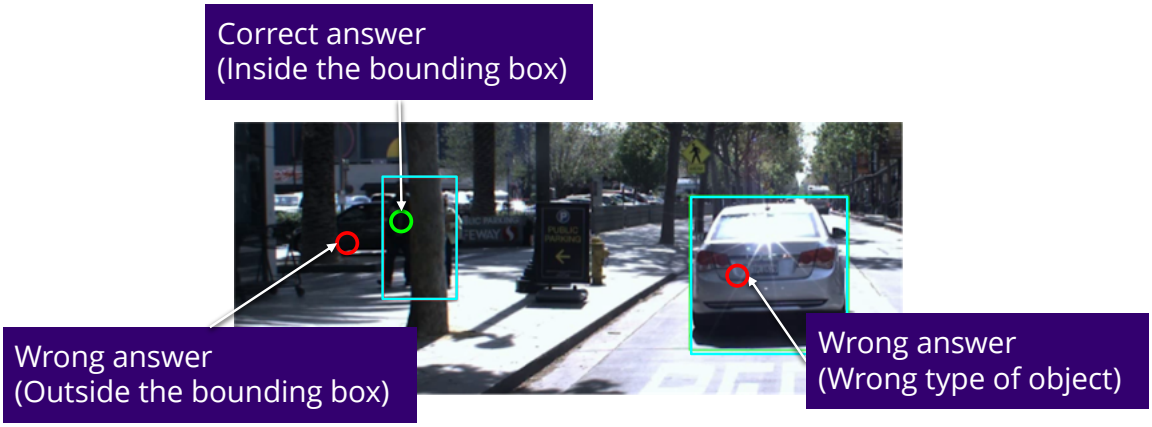


Figure 5.3: Illustration for labeling correct and incorrect anwers

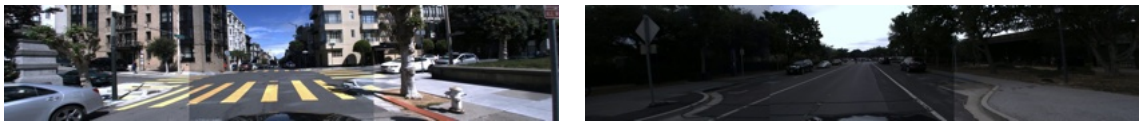
Independent Variables

- Duration (minutes): The accumulated driving time at each pause for the SAGAT question (Range (R): 0.7-17.8, Mean (M) = 8.81, Standard Deviation (SD) = 5.03).

The following measures were extracted for the scene at pause.

- Roadway type (3 levels): segment, intersection, and roundabouts.

- Lighting (2 levels): bright, less bright. The 'bright' scenario shows higher contrast and brightness compared to 'less bright' (Figure 5.4). Drivers may find it harder to identify objects under less bright road conditions. All scenarios were in the daytime. Darker scenarios were excluded from our eligible scenario pool since it is hard to detect objects even manually from the unclear dark scenarios. Including darker scenarios did not appear to yield a reasonable SA score under this study that is comparable with real-world driving.



(a) Bright (b) Less bright

Figure 5.4: Scenarios with different lighting conditions

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

- $P_{ped+cycle}$: Percentage of vulnerable road users, pedestrians, and cyclists, out of total objects, including OOIs and non-OOIs (R: 1-27, M = 5.15, SD = 5.24).

- P_{OOI}: Percentage of objects of interest; this includes pedestrians, cyclists, and vehicles (R: 1-15, M = 5.21, SD = 2.79).

- $P_{OOI,peripheral}$: Percentage of objects of interest in peripheral view out of total OOIs (R: 0-4, M = 0.39, SD = 0.82).

- 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).

5.2.1 Data Analysis Method

The data was used to model participants' SA scores given environmental features. There were 1200 SA scores (40 participants × 30 scenarios) collected from this study. The distribution of SA scores is shown in Figure 5.5. Of these, 754 SA scores equal zero; that is, no OOI locations were correctly identified in the scenario.

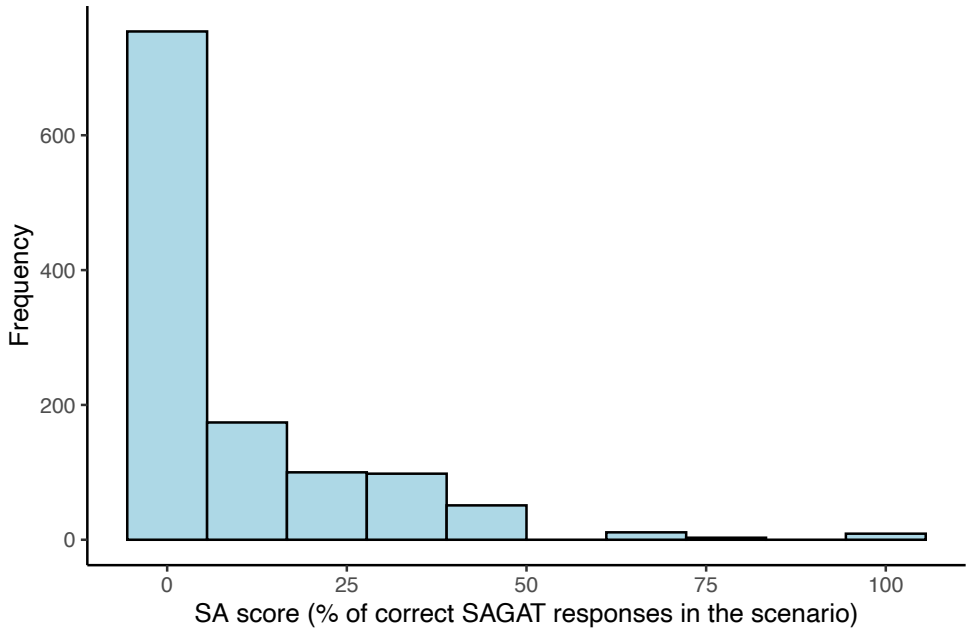


Figure 5.5: Distribution of SA scores

A hurdle model was used to account for excess zeros in a generalized linear model [17, 73]. The model consists of two parts: a mixed-effects binary logit model that shows the likelihood that the response outcome, SA score, was zero or not, and a truncated model that includes the positive outcomes when SA was not zero [69].

The mixed-effects binary logit model was used to predict the likelihood that the participant was able to detect an object correctly. An initial model was developed that considered all independent variables and interaction terms. Subject was included as a random effect. A stepwise approach was used to remove insignificant terms, which also included a review

of the goodness of fit using the Akaike Information Criterion (AIC).

The truncated portion or non-zero SA scores of the hurdle model was examined using a linear mixed-effects model [70]. The non-zero SA scores ranged from 6.67 to 100 (M=26.93, SD=17.12). The model included the independent variables described earlier, the interaction terms, and the subject (or participants) as a random effect. A stepwise approach was used to remove insignificant terms, which also included a review of the goodness of fit using AIC.

There were 30 responses per participant and significant differences were observed between participants. For that reason, the R package *lme4* was used to examine the mixed effects model within subjects as a random effect (random intercept) [4]. Interaction terms were also included in each model.

5.2.2 Results

Hurdle model - Mixed-effects binary logit portion

The final mixed-effects binary logit model is shown in Table 5.2. 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 will be zero or not. Scenarios with shorter duration, higher 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 5.6).

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. 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.

Table 5.2: Estimates of mixed-effects binary logit model predicting binary SA score

	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.: 754 AIC: 1490.7 Log Likelihood: -735.4 SD of Random effect: 0.54

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

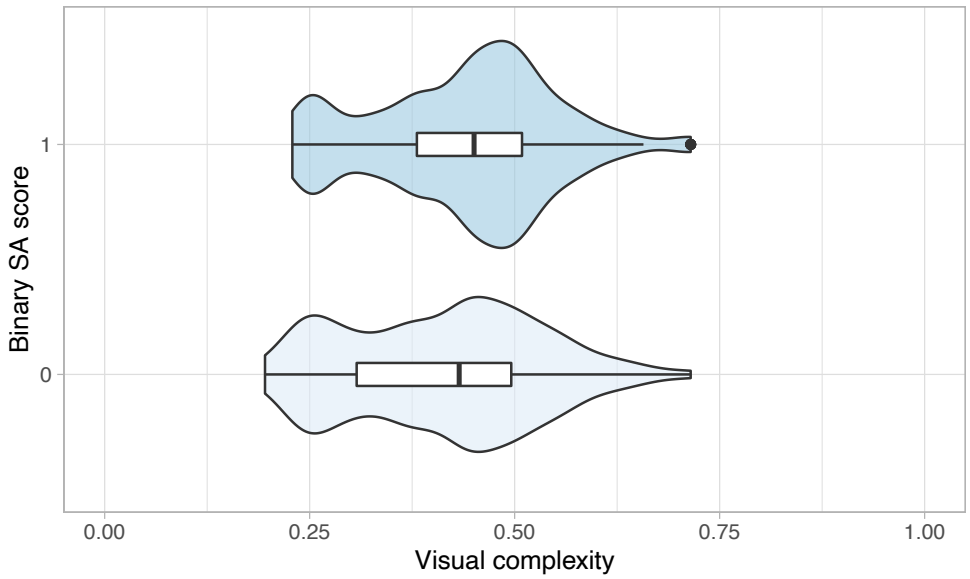


Figure 5.6: Violin plot of binary SA score given visual complexity

Hurdle model - Truncated - Linear mixed-effects portion

The result of the linear mixed-effects model is shown in Table 5.3. All the main effects except the Roadway type of segment had a significant impact on the SA score. Scenarios tend to have a greater SA score with lower visual complexity, fewer objects, smaller percentage of vulnerable road users (pedestrians and cyclists), smaller percentage of OOIs, smaller percentage of OOIs in the peripheral view, lighting condition of less bright compared to bright, and roadway type of roundabout compared to the intersection. Observations from a 2D heatmap (see Figure 5.7) showed that lower visual complexity (less than 0.5) generated higher SA scores (including some scores at 100% or close to 1).

In terms of the interaction terms, all were significant except two, which are the interaction term of visual complexity and N_{object} and the interaction term of visual complexity and roadway type(roadway). Each additional one percent of $P_{ped+cycle}$ increases the impact of P_{OOI} on SA score by $0.58 \times 0.01 = 0.0058$. Since this coefficient is positive, the more vulnerable users in the scene, the stronger impact of P_{OOI} will be and vice versa. Each additional one unit of N_{object} increases the impact of $P_{OOI,peripheral}$ on SA score by 0.03. Since this coefficient is positive, the more objects in the scene, the stronger impact of $P_{OOI,peripheral}$ will be and vice versa.

When the lighting condition is fixed as less bright, roadway type as segment will increase the SA score by 0.07 compared to intersections; roadway type as roundabout will decrease the SA score by 0.73; an additional 0.1 increase in visual complexity will increase SA score by 0.034.

When the roadway type is fixed as roadway segment, lighting condition of less bright will increase the SA score by 0.07 compared to bright condition. When the roadway type is fixed as roundabout, lighting condition of less bright will decrease the SA score by 0.73; additional 0.1 increase in visual complexity will decrease SA score by 0.848.

Table 5.3: Estimates of linear mixed-effects model predicting SA score

	Estimate	S.E.	t value	Pr(> |t|)	95% CI
Intercept	1.10 ***	0.11	9.81	<0.001	(0.88, 1.32)
Visual complexity	-0.64 **	0.24	-2.63	0.008	(-1.12, -0.16)
Lighting					
Bright					
Less bright	-0.18 **	0.06	-3.08	0.002	(-0.30, -0.07)
Roadway type					
Intersection					
Segment	-0.08	0.06	-1.26	0.207	(-0.20, 0.04)
Roundabout	3.32 ***	0.89	3.72	<0.001	(1.57, 5.07)
Object property					
N_{object}	-0.02 ***	0.01	-4.22	<0.001	(-0.03, -0.01)
$P_{ped+cycle}$	-0.27 ***	0.07	-4.04	<0.001	(-0.40, -0.14)
P_{OOI}	-0.75 ***	0.07	-11.02	<0.001	(-0.89, -0.62)
$P_{OOI,peripheral}$	-0.47 **	0.15	-3.20	0.001	(-0.76, -0.18)
Interaction terms					
Visual complexity:N_{object}	0.02	0.01	1.66	0.091	(-0.00, 0.04)
$P_{ped+cycle}$:P_{OOI}	0.58 ***	0.14	4.27	<0.001	(0.32, 0.85)
N_{object}:$P_{OOI,peripheral}$	0.03 ***	0.01	4.21	<0.001	(0.01, 0.04)
Lighting(less bright):Roadway(segment)	0.07 *	0.03	2.12	0.032	(0.01, 0.13)
Lighting(less bright):Roadway(roundabout)	-0.73 **	0.26	-2.85	0.006	(-1.23, -0.23)
Visual complexity:Roadway(segment)	0.08	0.13	0.61	0.518	(-0.17, 0.33)
Visual complexity:Roadway(roundabout)	-8.48 ***	2.31	-3.67	<0.001	(-13.01, -3.95)
Visual complexity:Lighting(less bright)	0.34 *	0.14	2.48	0.014	(0.07, 0.61)

Number of Obs.: 446 AIC: -521.6 SD of Random effect: 0.02 Conditional R^2: 0.53 Marginal R^2: 0.52
*** $p < 0.001$, ** $p < 0.01$, *$p < 0.05$

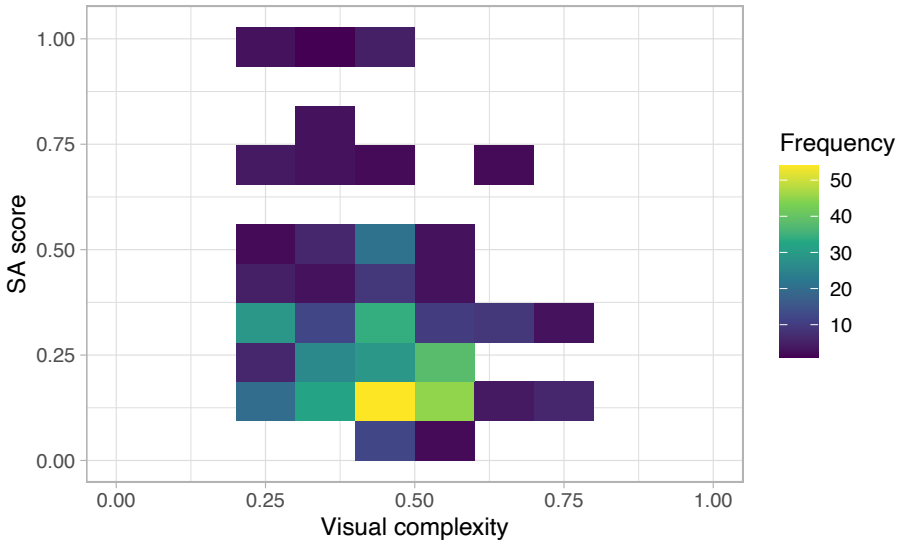


Figure 5.7: Heatmap of 2D bin counts for SA score given visual complexity

5.2.3 Findings

Results from the hurdle model showed that both higher visual complexity and more objects led to a significantly greater likelihood of getting non-zero SA scores for a scene. This indicates that in a condition with high visual complexity or more objects, it is easier for drivers to perceive some objects but will be harder for them to perceive a larger proportion of objects. This may relate to drivers' information processing mode in developing SA. Drivers alternate between "goal driven" and "data driven" information processing mode in SA development [34]. When there is too much information/input data from the environment, drivers will only react to the cues that are most salient. Greater visual complexity or more objects in the scene corresponds to a greater amount of visual information and clutter in scenarios [26]. This explains why drivers are more likely to perceive at least one object but find it difficult to perceive all objects of interest in complex environments. That is, higher visual complexity or more objects in the environment lead to a greater likelihood of getting non-zero SA scores but the SA scores tend to be lower.

The visual complexity and number of total objects consist of two information types: useful OOIs and useless objects. The useless objects only distracts 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 not only filter out useless information but also help drivers develop SA for missed OOIs.

In terms of the type of objects, the percent of vulnerable road users (i.e., pedestrians and cyclists) also has the same association as the three variables discussed above. That is, a higher percent of vulnerable road users leads to a greater likelihood of non-zero SA scores but the scores are lower. Since there are only three types of objects, the percent of cars will be $1 - P_{ped+cycle}$ and thus has an opposite effect when compared to the percent of vulnerable road users. In other words, a higher percentage of cars relates to a smaller likelihood of perceiving at least one object and a greater SA score, but the higher percentage of vulnerable road users relates to a greater likelihood of perceiving the object but a lower

SA score. Please note that a higher percentage of cars does not necessarily mean that the only object is a car since videos were selected from events with crossing pedestrians. This further emphasizes the need to help drivers develop SA, especially for vulnerable road users.

The primary goal was to create an algorithm that can predict drivers' situation awareness in a car. For that reason, an easy-to-compute measure "visual complexity" was selected to represent the visual clutter or the amount of visual information in a scene. Model results showed it is a reasonable variable to use. First, it has a weak correlation (< 0.3) with all measures used in this study (i.e., not correlated with other predictor variables). Second, although it has the same trend as the number of total objects in the scene, the coefficients of the interaction terms are different in the two model portions. When it comes to perceiving some objects (a non-zero SA score) visual complexity and number of objects weaken each other's impacts, but when it comes to perceiving more objects (a higher SA score) visual complexity and number of objects facilitate each other's impact. This indicates that it is a different measure reflecting both the useful and the useless information from the number of total objects.

For this study, a SAGAT method that is spatially continuous was proposed and examined. The results showed that a more precise SA measure can be obtained and future studies can be conducted to identify optimal SA scores for various environmental conditions. As an example, the distance from the actual object and the location marked by drivers can be a meaningful measure of SA. Additionally, future studies can consider how much each bounding box should be expanded to determine a correct and incorrect answer using the proposed SAGAT method. This may lead to more meaningful SA scores as the bounding box is associated with the calculation of the SA score. In general, the findings from this study provide a foundation for future studies on using SA in complex driving environments.

5.3 Summary

Our proposed method for collecting SA has the potential to yield quantitative measures for higher-level SA such as comprehension and prediction, as indicated by the results and brief discussion in each subsection. In addition, we found that environmental factors such as the number of objects, object size, visual complexity, and roadway type have a significant

impact on driver SA. Therefore, it would be reasonable to include these factors in our predictive model as we move forward.

Chapter 6

AIM 3: PREDICT DRIVERS' SA USING EYE-TRACKING DATA

This chapter introduces the eye gaze metrics used as predictors and how they have been extracted from the eye-tracking data. Predictive methods of SA based on eye gaze metrics, environmental, object, and driver features are presented and compared.

Both the recognized location and the type of the objects were recorded, which is a two-phase procedure, as depicted in Figure 6.1: 1) object localization, 2) delve into more details - object recognition. The methods in this chapter will also follow this hierarchical structure, i.e., each method contains two phases. Binary classification models predicting whether the object was correctly located will be fitted and compared, followed by another binary classification model predicting whether the object type was correctly recognized. The two phases of each method share the same pool of predictors.

The chapter unfolds as follows: Section 6.1 delves into the processing of eye-tracking data, while Section 6.2 introduces the SA responses of each phase of the two-phase methods. Section 6.3 elaborates on the feature sets employed in the predictive models, followed by a description of the data-splitting process in Section 6.4. Section 6.6 outlines a calibration method attempted but ultimately abandoned. Section 6.7 encompasses the baseline methods, proposed approaches, binary classification models, performance metrics, and model comparisons. Finally, Section 6.8 provides a brief discussion of the obtained results.

6.1 Eye-Tracking Data Processing

The object trajectories have been estimated in preliminary work for a one-second window before the SAGAT pause. Eye-tracking data will also be processed for this one-second window T in order to be mapped with the estimated object bounding boxes along the trajectory.

The raw eye-tracking data has been recorded at 30 Hz and has been synchronized and

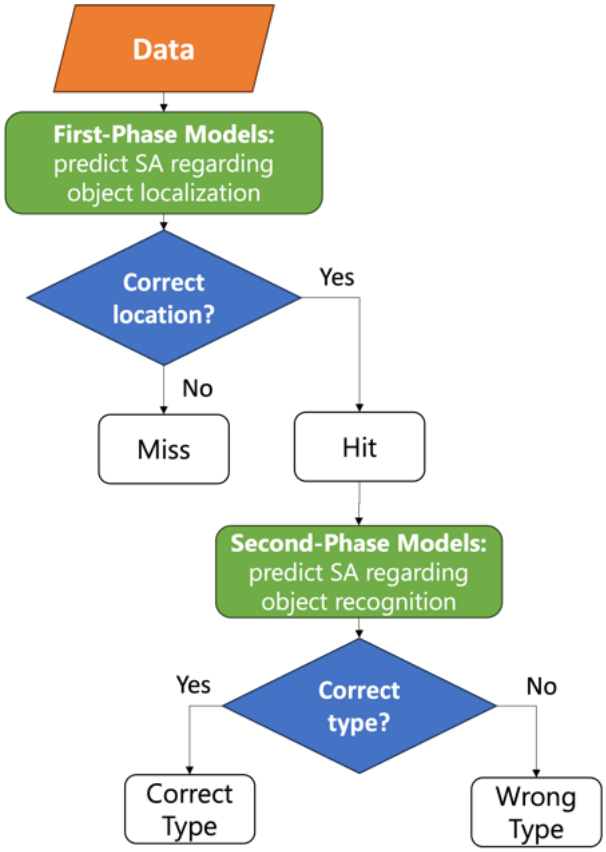


Figure 6.1: Two-phase predictive model

projected to the scenario video's coordinate system. In this case, the eye tracking data and the SA responses as x-y coordinates are now in the same coordinate system, which makes the computation of distances between the gaze points and the objects possible.

Before describing the detailed features, there is one thing to be noted. The coordinate system of the scenario video file is a Cartesian coordinate system with the top left corner as the reference point as shown in Figure 6.2(b), however, in the actual simulator setting, there are angles 2β between adjacent displays. In this case, the distance between two points on different displays computed in this Cartesian coordinate system will not reflect their actual distance in the real 3D setting. Therefore, a polar coordinate system as depicted in Figure 6.2(c) was established with the participant as the reference point. Since the distance between each display and the participant is equal, we assume the three-channel display

forms an arc with a radius of $r = 48''$.

The parameter, the angle of a unit arc (length = 1 px, radius = $48''$) corresponds to, needs to be calculated in order to convert the distance between two points in the Cartesian coordinate system to the polar coordinate system. A correspondence was detected between the width of a single-channel video (1920 px) and an arc with an angle of 2α so that α is calculated as:

$$\tan(\alpha) = \frac{w/2}{r} \Rightarrow \alpha = 23.479^\circ \tag{6.1}$$

Hence, the parameter equals:

$$2\alpha \sim 1920 \text{ px} \Rightarrow 1 \text{ px} \sim 0.0244572^\circ \tag{6.2}$$

The data processing procedures have been coded using R and tested on a small set of data. Columns of this processed data are described in Table 6.1. Note that only gazes that are labeled as fixations are used when processing metrics.

Table 6.1: Eye Tracking Data Codebook

	Column name	Description
1	tobiiTS	Tobii eye tracker's device time, continuous
2	sysT	System time of the computer playing the scenario video, continuous
3	vidT	Video elapsed time, will not increase when paused
4	last_event	Labeling the start and end of each scenario and SAGAT pause
5	scene_frame_num	Scenario video frame index
6	tobii_frame_num	Tobii video frame index
7	Pupil_left	Pupil size, left
8	Pupil_right	Pupil size, right
9	Is_fixation	Tobii's IBT algorithm, below 30 is saccade, above is fixation, degree/sec
10	Gaze_Velocity	Gaze point moving velocity, degree/sec
11	SceneCamFile	Scenario video filename
12	TobiiCamFile	Tobii video filename
13	GazePointInTobii_x	x coordinate in Tobii eye tracker video, range(0-1)
14	GazePointInTobii_y	y coordinate in Tobii eye tracker video, range(0-1)
15	GazePointInScene_x	x coordinate in scenario video, range(0-1)
16	GazePointInScene_y	y coordinate in scenario video, range(0-1)

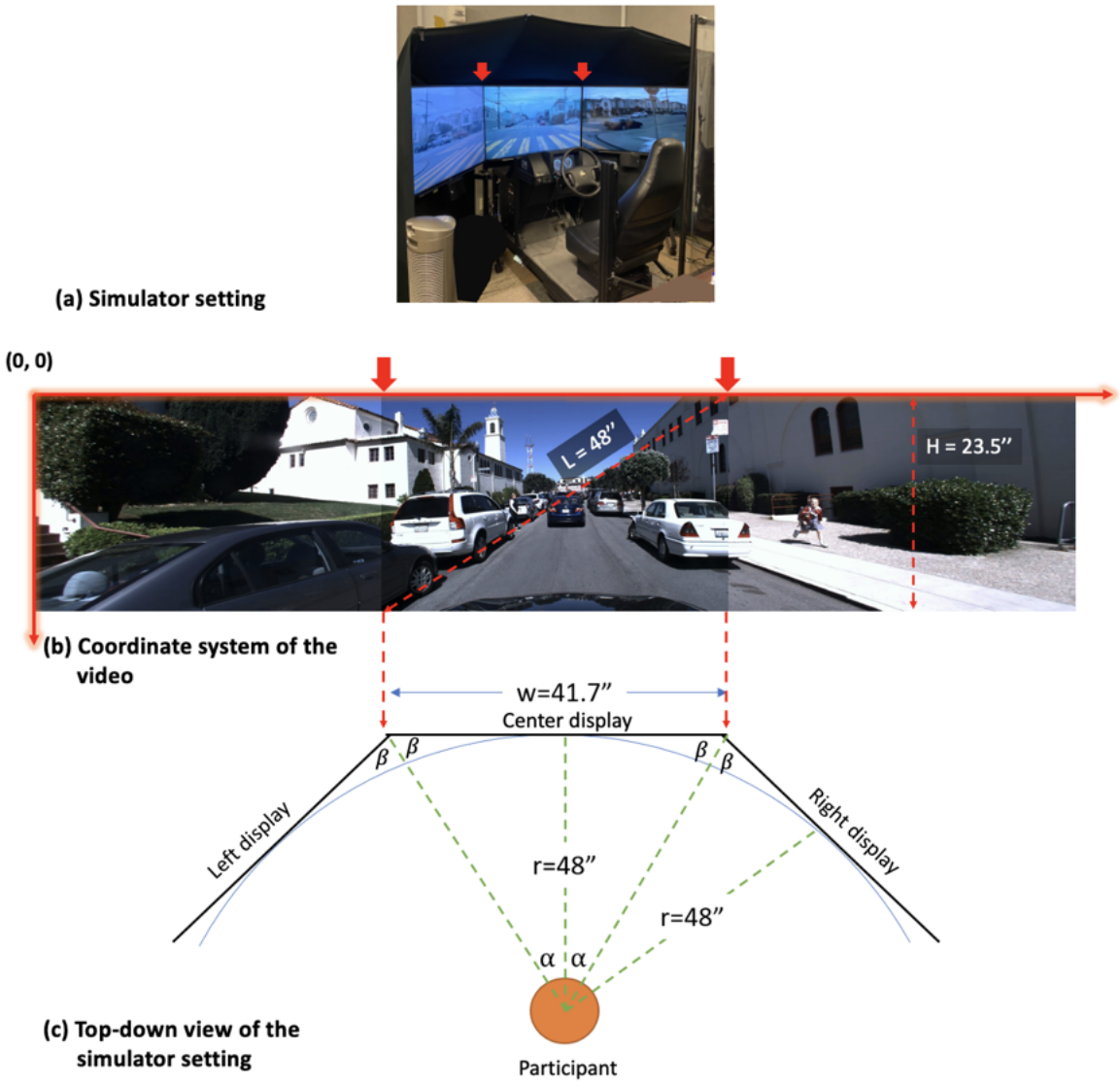


Figure 6.2: Simulator setting and key geometric parameters

6.2 Response

Object-wise SA will be the focus of the proposed work considering our research goal to selectively provide assistance. After data cleaning, 6192 data points each representing one object for one participant, were used for modeling in this chapter.

6.2.1 SA regarding Object Localization

The first-phase models predict the situation awareness regarding object localization by determining whether the object's location has been accurately identified. The SA response indicating object localization is a binary variable for the first model as depicted in Figure 6.1.

- *Hit* (21.2%) - the location of the object has been correctly identified. If any touch has been detected within the object's bounding box area or its trajectory, the object's location.
- *Miss* (78.8%) - the object was not detected by the participant. No touch located within the object's bounding box or its trajectory has been detected.

6.2.2 SA regarding Object Recognition

The second-phase models predict the situation awareness regarding object recognition by determining whether the object's type has been accurately identified. The data set for this phase has 1315 data points. The SA response indicating object recognition is a binary variable for the second model as depicted in Figure 6.1.

- *Correct Type* (78.6%) - both the object location and its type have been correctly identified.
- *Wrong Type* (21.4%) - the location of the object was accurately recognized, but its type was misidentified.

6.3 Predictors

6.3.1 Gaze Enclosure

Gaze enclosure is a binary indicator that denotes whether a gaze point has been detected within the bounding box during the specified time window.

6.3.2 Gaze Point-Based Features

(a) Distance

Gaze-point-based features utilize both the spatial alignment between target objects and participants' gaze point coordinates, as well as the velocity of gaze points. Two types of object-gaze point distance have been considered:

- Distance between the gaze point and its closest point to the bounding box.
- Distance between the gaze point and the center of the bounding box.

Humans tend to perceive objects as whole entities rather than focusing on gaps within them. This inclination to mentally complete shapes and figures is referred to as "closure". The law of closure dictates that individuals perceive objects, such as shapes, letters, or pictures, as complete even when they are not fully formed. In essence, when certain parts of a whole picture are missing, our perception automatically fills in the visual gap [9]. Thus, in our case, it is reasonable to use the distance between eye fixations and specific points of the object to indicate how close the drivers were looking at the object.

Three features were extracted for each type of object-fixation distance as follows.

- GD_{pause} - gaze point distance to an object at a pause, measured in degree.
- GD_{min} - minimum gaze point distance to an object within the analysis time window T, measured in degree.
- GD_{avg} - average gaze point distance with an object during the analysis time window T, measured in degree.

(b) Velocity

- GV_{mean} - average gaze point moving velocity within in T, measured in degrees per second.
- GV_{pause} - gaze point moving velocity at pause, measured in degrees per second.

The velocity of gaze point movement at minimum distances was initially considered but ultimately excluded due to a high percentage of missing data.

6.3.3 Visual Sensory Dependent Features

There are two main parts in the human visual sensory system: foveal vision and peripheral vision. Foveal vision is used for scrutinizing highly detailed objects, whereas peripheral vision is used for organizing the broad spatial scene and for seeing large objects. In the driving context, foveal vision is optimized for sensitivity in object detection and peripheral vision is good at the recognition of motion. Therefore, it is important to involve this set of features in the predictive model.

This study presumes that drivers are using non-foveal vision for tracking detected objects and detecting moving objects. Based on clinical research [49], four vision sensory radius ranges will be considered: fovea ($\theta_1 = 2.5$ degree), parafovea ($\theta_2 = 4.1$ degree), perifovea ($\theta_3 = 9.1$ degree) and macula ($\theta_4 = 15.0$ degree). Three features will be extracted for each radius range. For radius range i the three features will be extracted using two types of object-gaze distance:

- $VS^{(i)}_{elapse}$ - time elapsed since the last frame where the gaze distance is less than θ_i counted at pause. If no fixation is within the range, the analysis window T will be used.
- $VS^{(i)}_{dwell}$ - total dwell time measured in seconds when the gaze distance is less than radius θ_i.
- $VS^{(i)}_{avg}$ - average gaze distance measured in degree while the distance is within a radius threshold of θ_i. For gaze distance always larger than θ_i, the average distance during the whole analysis window T will be used.

6.3.4 Environmental, Object and Demographic Features

(a) Environmental Features

- OE_{road} - roadway type, a binary indicator of whether the roadway type is segment or intersection.
- OE_{objcnt} - object count, represents the number of objects in the scenario.
- OE_{pedpct} - the percentage of the number of pedestrians out of the number of all objects in the scene at pause.
- OE_{light} - lighting condition, a binary indicator of whether it was bright or less bright.
- OE_{viscpx} - visual complexity score of the scenario, computed using equation 4.1 as the ratio between the compressed and uncompressed at-pause image file size. It is a continuous variable ranging from 0 to 1 with a higher value indicating a greater visual complexity of the scene.

(b) Object Features

- OO_{type} - object type, a three-level categorical variable indicating the type of the object of interest, including pedestrian, car, and cyclist.
- OO_{size} - object size, a continuous variable indicating the size of the object of interest, measured in pixels.

(c) Demographic Features

- OD_{age} - age group, a binary indicator reflecting whether the driver was in the 40- or 40+ age group.
- OD_{gen} - driver's gender, a binary indicator with two levels male and female.
- OD_{drvfrq} - driving frequency, represented by three levels including less than weekly, at least weekly, and at least daily.

6.4 Data Split

As indicated by the average SA accuracy across scenarios and drivers shown in Figure 6.3, SA accuracy exhibits slightly greater variability across scenarios compared to across drivers. In real-world driving scenarios, it is more pragmatic to generalize the model's applicability to various driving situations. The optimal model is one that has been exposed to diverse operators and limited scenarios, yet can achieve better predictions for unobserved situations.

Out of the 75 scenarios, 15 (20%) have been randomly assigned to the test set, while the remaining 60 (80%) are assigned to the training set. This allocation results in a training set consisting of 4921 rows and a test set of 1271 rows for object localization, and 1127 rows for training and 188 rows for testing for object recognition.

Notably, since each of the 40 drivers has experimented in 30 distinct scenarios, both the training and test sets encompass data from each driver.

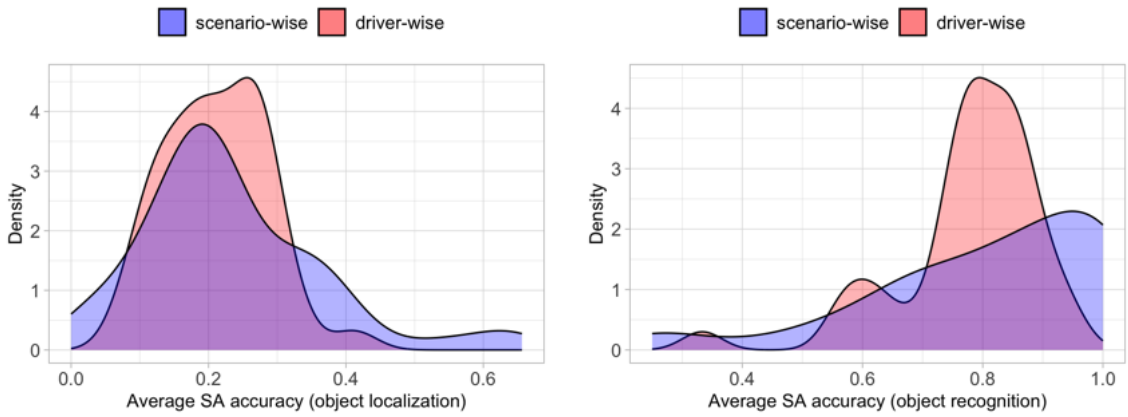


Figure 6.3: Average SA accuracy distribution across scenarios and drivers

6.5 Subsampling

In classification tasks, addressing class imbalances is crucial for effective model fitting. Three types of sampling methods have been developed for this purpose:

- Downsampling/undersampling: This method involves randomly subsetting all classes in the training data to match the frequency of the least prevalent class.

- Upsampling/oversampling: Here, the minority class is randomly sampled (with replacement) to match the size of the majority class.

- Hybrid methods: These approaches combine aspects of both up and downsampling to incorporate the strengths from both.

Two prominent techniques of the hybrid methods are the synthetic minority oversampling technique (SMOTE) [14], which down-samples the majority class while generating synthetic data points in the minority class; and random over-sampling examples (ROSE) [68], which is a bootstrap method that helps with binary classification when dealing with rare classes. Both work with continuous and categorical data by creating synthetic examples based on a conditional density estimate of the two classes. This study utilized the *DMwR* package (Version 0.4.1) for SMOTE subsampling and *ROSE* package (Version 0.0-4) on each training set with different predictor sets.

6.6 Calibration

Model calibration aims to ensure that the estimated class probabilities are consistent with what would naturally occur. A model with poor calibration indicates low reliability, and we might be able to post-process the original predictions to coerce them to have better properties following the procedures depicted in Figure 6.4.

After employing subsampling techniques like SMOTE and ROSE, sometimes it's essential to refine the predicted probabilities of the classification model using calibration methods depending on the output results. Synthetic examples generated by the subsampling methods may introduce noise into the class distribution. Therefore, selecting calibration methods capable of handling noisy or imbalanced data is helpful. Isotonic regression, a non-parametric method, can offer greater robustness to noisy data and may serve as a suitable choice post-subsampling. This study utilized the *isoreg* function from package *stats* for the isotonic regression.

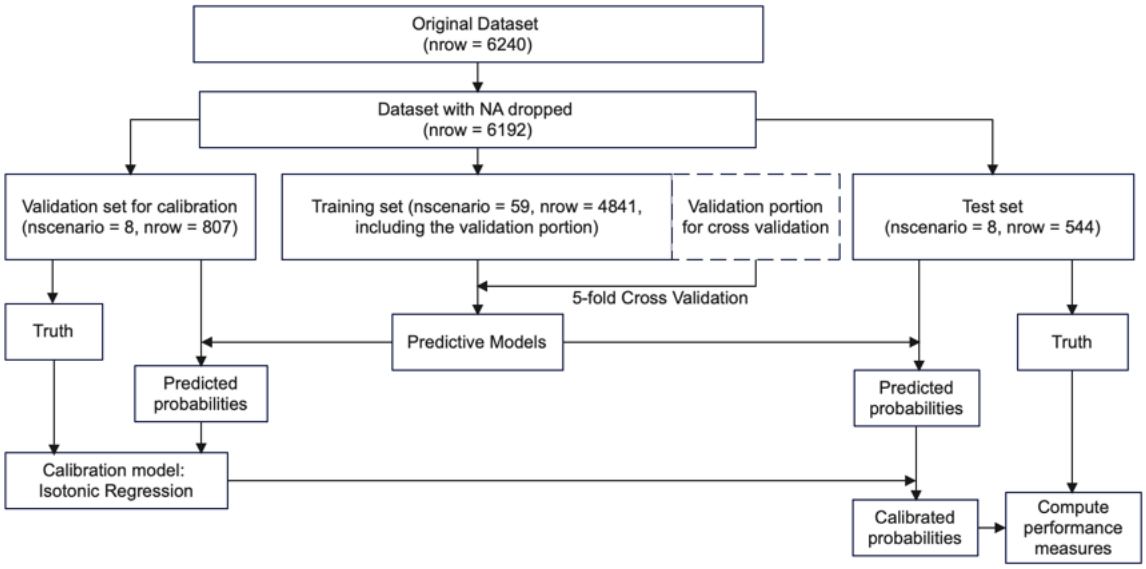


Figure 6.4: Flowchart for calibration process

When there is the need to calibrate the output probabilities, among the 75 scenarios, a validation set containing 8 scenarios (10%) is split out in addition to the training test with 59 scenarios (80%) and the test set with 8 scenarios (10%). Separating a validation set for fitting the calibration model aims to avoid touching the data in the test set during calibrating or reusing the data in the training set which may cause overfitting.

A calibration plot is a good way to diagnose the calibration need of model results. It is a graphical representation that illustrates the agreement between predicted probabilities or scores from a model and the observed outcomes. It helps assess the reliability of a predictive model by showing how well the predicted probabilities align with the actual probabilities of events occurring.

Figure 6.5 shows some examples of calibration plots. The closer the line is to the diagonal, the more reliable the model is. R function *cal_plot_windowed()* from package *probably* (Version 1.0.3) was utilized for drawing calibration plots. We will only focus on the top right plot, and the two at the bottom to explain the plots. The top right plot shows the results of fitting the logistic regression model for Method Baseline 3 (methods are described in more detail in the following section), which has a big drop of around 0.7,

meaning that when the predicted probabilities are around 0.7 to 0.75, the actual event rate of being 1 rather than 0 is way less than 0.7 and 0.75, indicating a poor calibration. The bottom right plots show the results of fitting logistic regression based on the subsampling examples using SMOTE. The range of predicted probabilities becomes wider, and the drop becomes less acute but it is much further from the diagonal. Calibration is helpful in this case, the results after calibrating the probabilities using the isotonic regression are depicted in the bottom right plot. The line comes closer to the diagonal, indicating an improved calibration or reliability of the model.

Unfortunately, upon initial inspection of the output results, it appears that following calibration, all data points were predicted as zero across all methods. This outcome stems from the relatively small validation set, resulting in a poorly fitted isotonic regression model that assigns all predictions to a single category. Consequently, our data does not effectively support calibration following subsampling, and no calibration was implemented for any model presented in this dissertation.

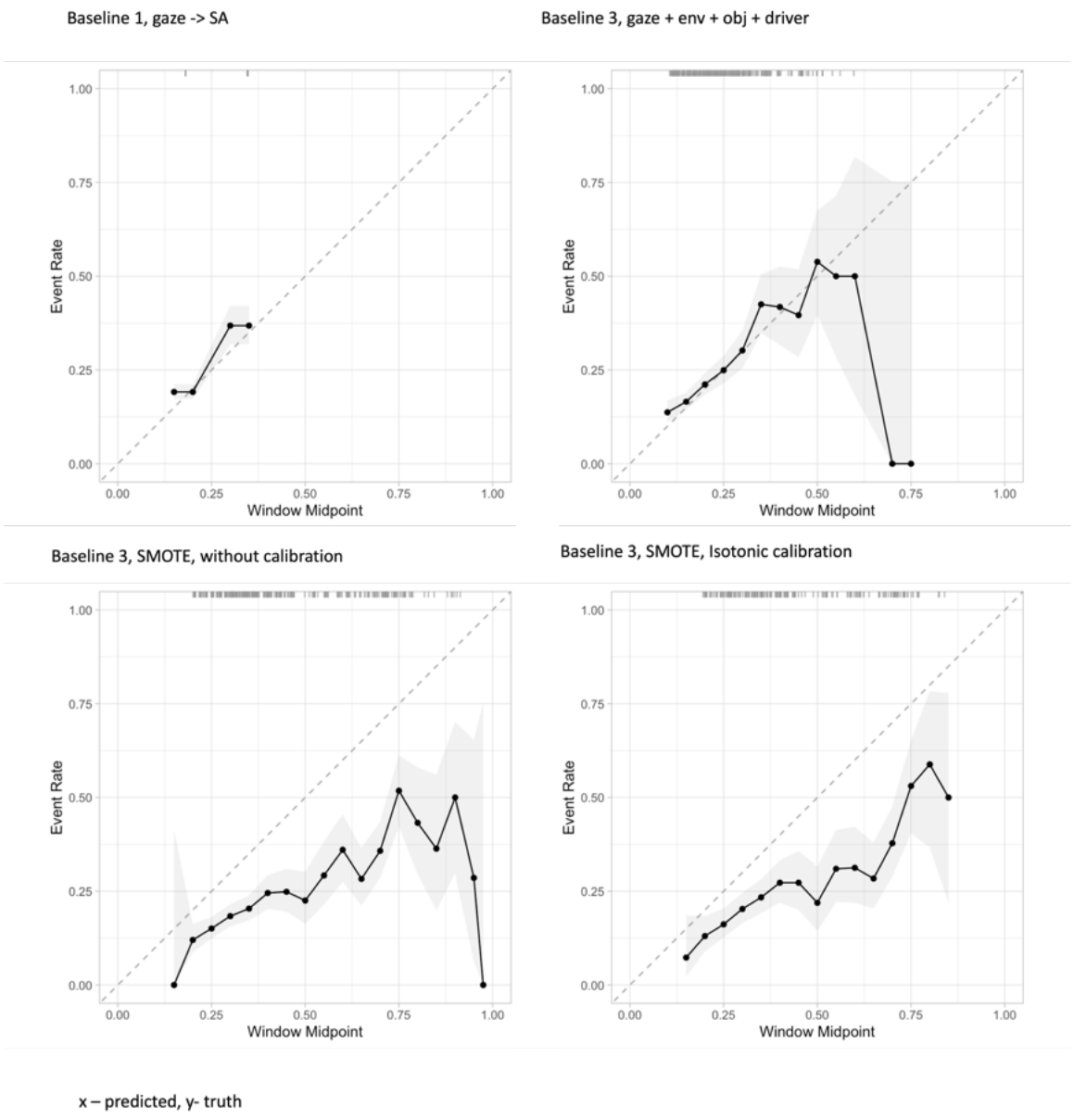


Figure 6.5: Calibration plot examples

6.7 Method Evaluation

The same predictor set will be utilized for both phases of the two-phase method set. In the first phase, the model predicts SA concerning object localization, while in the second phase, it predicts SA regarding object recognition. A summary of the feature sets for all

methods is provided in Table 6.2, with detailed descriptions available in the subsequent subsections. The numbers within the cells represent the quantity of predictors employed in each respective method. Additionally, numbers within brackets alongside "PCA" indicate the application of principal component analysis (PCA), with the enclosed number denoting the count of PCA-transformed predictors utilized in the method.

Table 6.2: Feature Sets of Each Method for Aim 3

Method	Model	Gaze in bbox	Gaze point-based features	Visual sensory dependent features	Environmental features	Object features	Driver features	Predictors count
Baseline1	rule-based	1						1
Baseline2	learning	1						1
Baseline3	learning	1			5	2	3	11
M1	learning		2 (PCA) + 2		5	2	3	14
M2	learning			6 (PCA)	5	2	3	16
M3	learning		7 (PCA) + 2		5	2	3	19

6.7.1 Baseline Methods

Baseline 1: Gaze enclosure rule-based

The underlying assumption is that "looking implies seeing," meaning that the direction of a driver's eye gaze directly indicates whether they have perceived the object. This approach relies solely on whether the gaze point is located within the bounding box to predict the driver's situational awareness. The predictor "gaze enclosure" is used as the only predictor to predict driver SA with a rule-based model.

Baseline 2: Gaze enclosure learning-based

The fundamental assumption aligns with Baseline 1 that the direction of a driver's gaze directly indicates whether they have perceived the object. However, in this approach, learning-based models are employed. These models are consistent with those utilized in other methods, and the binary classification models are detailed in Section 6.7.3.

Baseline 3: Baseline 2 + environmental, object and driver features

This method echoes the assumptions of the previous two. The environmental, object, and driver features investigated in Aim 2 were added, and learning-based models were employed.

6.7.2 Proposed Methods

Method 1: Effect of incorporating gaze point-based features

In the first proposed method, the binary predictor "gaze enclosure" is replaced by gaze point-based features. A notable correlation has been observed between all pairs of distance-related gaze-point-based features, regardless of whether they were computed using the object's center point or the nearest point on the bounding box's perimeter. To retain information from each predictor and address the issue of multicollinearity, PCA was applied to the six predictors (two distance types multiplied by three features per distance type). Note that only the training set was utilized for fitting PCA. Subsequently, the PCA model fitted on the training set was applied to the test set, ensuring that the data in the test set was not accessed prior to testing.

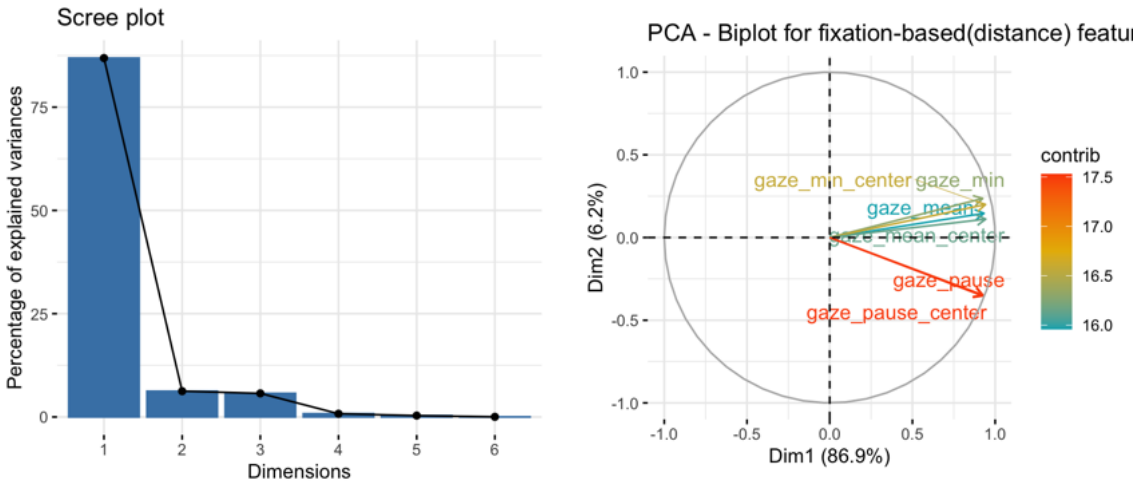


Figure 6.6: Scree plot (left) and biplot (right) of PCA results for distance-related gaze-point-based features

The first two components were selected as they accounted for 93.1% of the variance, surpassing our preset threshold of 90%, which has been used throughout the study. The scree plot depicted in Figure 6.6 illustrates the percentage of variance explained by each principal component, while the biplot on the right demonstrates the influence of each predictor on the first two principal components.

Method 2: Effect of incorporating visual sensory dependent features

In the second proposed method, gaze-point-based features from Method 1 are replaced with visual sensory-dependent features. Similar to the first method, multicollinearity arises, prompting the utilization of PCA to address this issue. Six principal components were selected, explaining 90.5% of the variances of the 24 predictors, comprised of four visual sensory radius ranges multiplied by three features per distance type. The scree plot depicted in Figure 6.7 illustrates the percentage of variance explained by each principal component, while the biplot on the right demonstrates the influence of each predictor on the first two principal components.

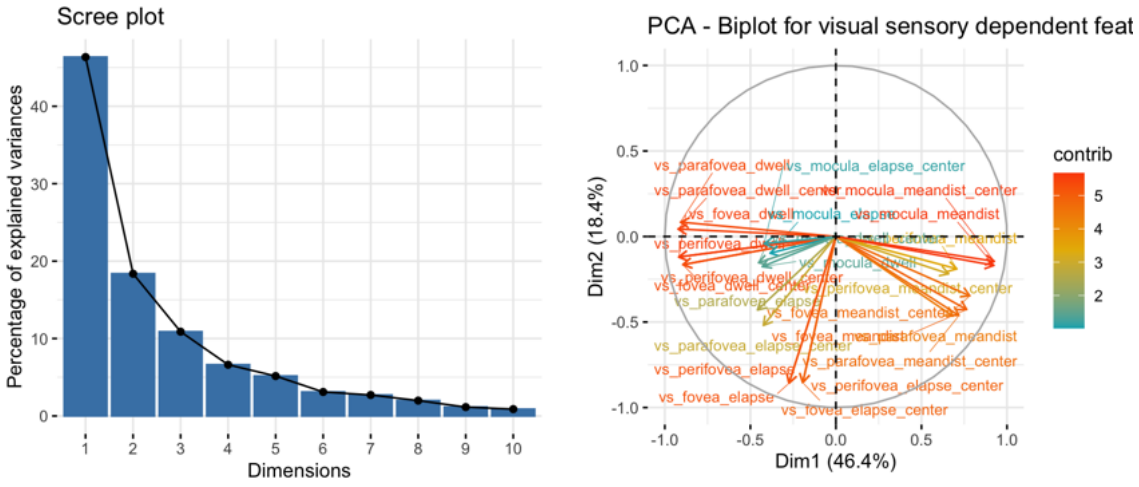


Figure 6.7: Scree plot (left) and biplot (right) of PCA results for visual sensory-dependent features

Method 3: Effect of incorporating both gaze-point-based & visual sensory dependent features

The third proposed method incorporates both the gaze-point-based features and the visual sensory-dependent features. Strong correlations have been detected between the distance-related gaze-point-based predictors and the visual sensory dependent features, as depicted in Figure 6.9. Therefore, a PCA on all those features except gaze-point velocity features. The scree plot depicted in Figure 6.8 illustrates the percentage of variance explained by each principal component, while the biplot on the right demonstrates the influence of each predictor on the first two principal components. Seven principal components were selected, explaining 91.9% of the variances of 30 predictors.

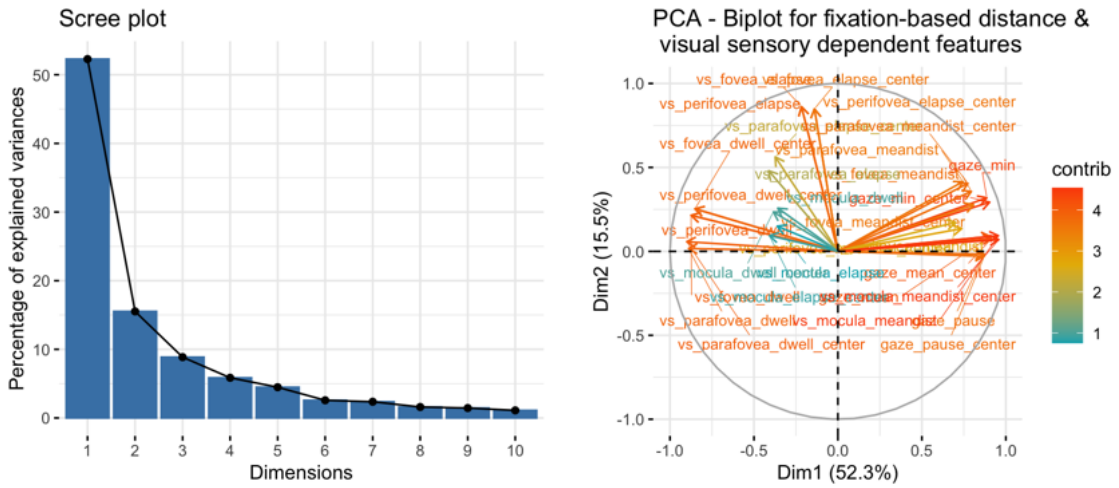
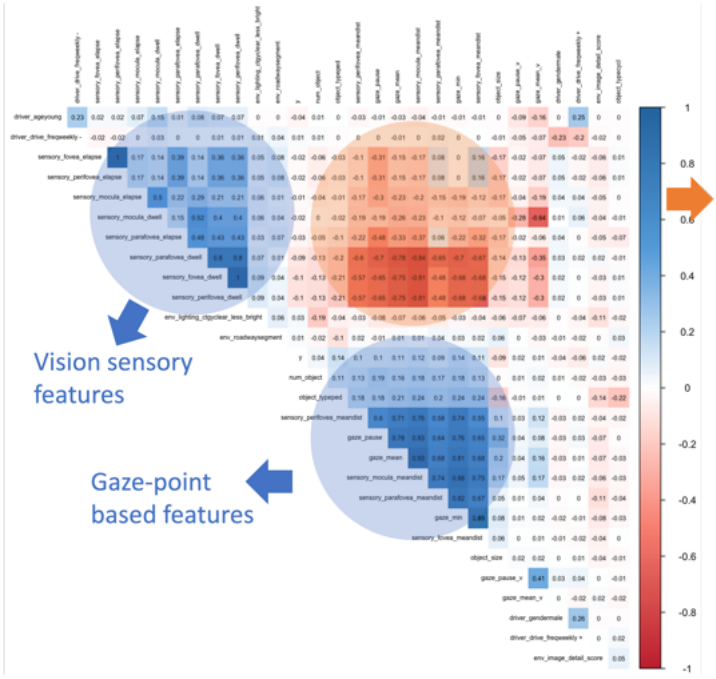
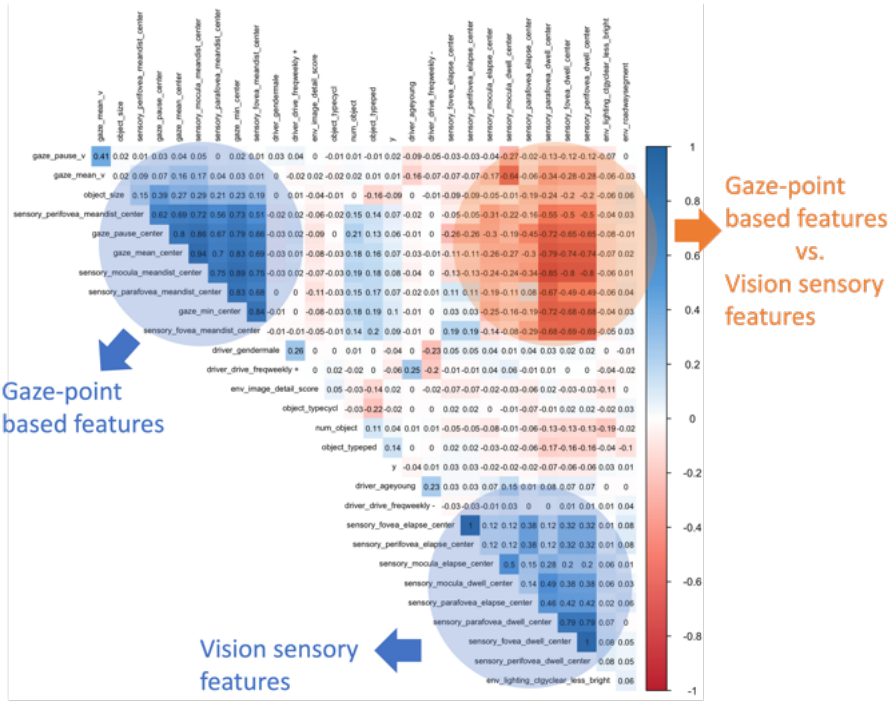


Figure 6.8: Scree plot (left) and biplot (right) of PCA results for distance-related gaze-point-based features and visual sensory-dependent features



*Gaze distance calculated as between the gaze point and its closest point to the bounding box.



*Gaze distance calculated as between the gaze point and the center of the bounding box.

Figure 6.9: Caption

6.7.3 Binary Classification Models

Binary classifiers will be employed for both phases of each method. The predictor sets will be the same for two phases of one method, but the binary classifiers can be different. Six learning-based classic binary classifiers have been employed: logistic regression, decision tree (entropy, Gini), random forest, and XGBoost linear and XGBoost tree. R package *caret* (Version 6.0-94) was used for training the models. All numerical predictors have been centered and scaled pre-training. A five-fold cross-validation was employed for all training processes.

6.7.4 Model Performance Measures

Performance computed for models are listed below:

- *Confusion Matrix* is a 2 × 2 matrix containing four cells, which are counts of true positive (TP), true negative (TN), false positive (FP), and false negative (FN).

- *Precision* is computed as the ratio of correctly predicted positive observations to the total predicted positive observations: $\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$

- *Recall* is defined as the ratio of correctly predicted positive observations to all observations in actual class: $\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$

- *Accuracy* is the proportion of correctly classified instances:
  $\text{Accuracy} = (\text{TP}+\text{TN})/(\text{TP} + \text{TN} + \text{FP} + \text{FN})$

- *Balanced Accuracy* is a performance metric used in classification tasks, especially when there is an imbalance in the class distribution. It is defined as the arithmetic mean of sensitivity (TP rate) and specificity (TN rate). A classifier that always predicts negative will have a balanced accuracy equal to 50%. Considering the data imbalance in this study, it is reasonable to compare the balanced accuracy results with 50%.
  $\text{Balanced Accuracy} = (\text{TP}/(\text{TP} + \text{FN}) + \text{TN}/(\text{TN} + \text{FP}))/2$

- *F1 Score* is a metric used to evaluate the performance of a classification model, particularly in tasks where the class distribution is imbalanced. It is the harmonic mean of precision and recall. The F1 score ranges from 0 to 1, where a higher score indicates better model performance. It provides a balance between precision and recall, making it a useful metric for assessing a model's ability to correctly classify both positive and negative instances, especially when there is an imbalance between the classes: $\text{F1 Score} = 2 \times (\text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall})$

- *AUC* stands for "Area Under the ROC Curve." It is a popular performance metric used to evaluate the performance of binary classification models.

  The ROC (Receiver Operating Characteristic) curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. It is created by plotting the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings.

  The AUC provides an aggregate measure of performance across all possible classification thresholds. A higher AUC value indicates better overall performance of the model in distinguishing between the two classes. The AUC ranges from 0 to 1, where AUC = 0.5 indicates that the model has no discrimination ability (equivalent to random guessing); AUC = 1 indicates perfect discrimination, where the model perfectly separates the positive and negative instances. In summary, AUC is a useful metric for evaluating the discrimination ability of binary classification models, particularly in scenarios with imbalanced class distributions or when the true positive and false positive rates are of interest across various threshold settings.

- *Brier Score* also known as the Brier loss or mean squared error, is a performance metric commonly used to assess the accuracy of probabilistic predictions made by a classification model.

  It measures the mean squared difference between predicted probabilities and the actual outcome for each instance in the dataset. Specifically, for a binary classification

problem, the Brier score is calculated as the average squared difference between the predicted probability assigned to the positive class and the actual outcome (1 for the positive class, 0 for the negative class). The formula for the Brier score is:

$$\text{Brier score} = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$$

Where N is the total number of instances in the dataset. f_i is the predicted probability assigned to the positive class for instance i. o_i is the actual outcome for instance i, where $o_i = 1$ if the instance belongs to the positive class and $o_i = 0$ if it belongs to the negative class.

The Brier score ranges from 0 to 1, where a lower score indicates better calibration and accuracy of the model's predicted probabilities. A Brier score of 0 indicates perfect accuracy, meaning the predicted probabilities perfectly match the actual outcomes, while a score of 1 indicates complete inaccuracy, equivalent to random guessing.

- *Performance Score* is a self-defined score computed as the average of F1 score and balanced accuracy. The reason for defining this score is articulated in the next paragraphs.

Considering the data's imbalance, we meticulously selected performance measures for the models. Specifically, we chose four metrics for model comparison. We then filtered models that performed above the 0.85 quantile in terms of both the Brier score and AUC. Subsequently, we arranged them by a performance score calculated as the average of the F1 score and balanced accuracy. This method yielded a model showcasing robust performance across all criteria. The top five models in the rank for each method phase were consequently chosen as the overall best models.

In comparing the best model for each method, we filtered models performed above the 0.5 quantile concerning the Brier score and AUC. Since we grouped the model results by method, with each group containing 18 models, significantly fewer than the total of 91 models, we adjusted the quantile to include more models in the final pool for performance

score comparison. We then sorted them by a performance score computed as the average of the F1 score and balanced accuracy.

6.7.5 Results

Altogether, one rule-based model was executed for Baseline 1. For the remaining methods, including Baseline 2 and 3, Methods 1 to 3, a total of six binary classification models were trained and tested using non-subsampled data, SMOTE subsampled examples, and ROSE subsampled examples, resulting in a combination of six times three times five, totaling 90 + 1 = 91 models for each phase of the method. Detailed performance results for each model can be found in Appendix A.1.

(1) Object Localization

The four primary performance metrics of all methods except Baseline 1 are illustrated in Figure 6.10. Notably, we adjusted the Brier score to "1 - Brier score" to maintain consistency across the measures, ensuring that a value closer to 1 signifies better model performance.

Models utilizing non-subsampled data and those employing SMOTE subsampled examples demonstrate better performance compared to ROSE. Conversely, Baseline 2 exhibits the poorest overall performance. Among the remaining methods, Method 3 outperforms the others in terms of balanced accuracy, while Baseline 3 exhibits the lowest performance in this regard. However, Method 1 and Method 2 display no discernible differences based on the plots. Furthermore, no significant disparities have been observed in terms of AUC, Brier score, and F1 score among the four methods, except for Baseline 2.

The top five models, ranked by their performance scores after filtering out those that exceeded the 0.85 quantile for both the Brier score and AUC, are presented in Table 6.3. When comparing balanced accuracy, where there's a baseline of 50% from always predicting negative, the model achieving the highest balanced accuracy is 69.2%. Regarding AUC, a range of 0.7 to 0.8 is typically considered acceptable [66]. Additionally, we are not displaying the best balanced accuracy or discussing the overall best balanced accuracy. Models that performed poorly in terms of AUC and Brier score have been filtered out, which may

performs well in terms of balanced accuracy at the cost of a poor performance for the rest. It is noteworthy that all the models in the top five are based on non-subsampled data; and that XGBTree and XGBLinear models perform best.

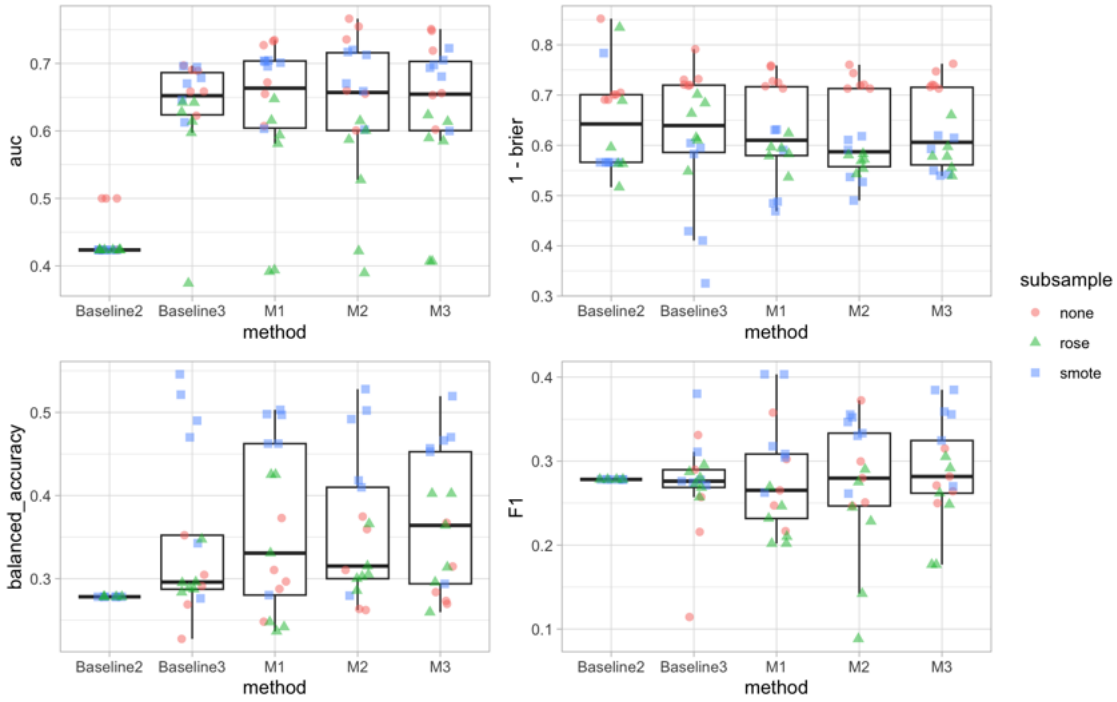


Figure 6.10: Aim 3 - Performance measures for object localization methods

Table 6.3: Aim 3 - Top Five Performance Results of All Predictive Methods for Object Localization

m.	ss.	model	TP	FP	FN	TN	acc.	b-acc.	F1	prc.	rec.	AUC	Brier	prfm.
M1	none	xgbTree	56	132	69	1014	0.842	0.666	0.358	0.298	0.448	0.733	0.243	0.512
M2	none	xgbLinear	65	123	96	987	0.828	0.646	0.372	0.346	0.404	0.755	0.240	0.509
M3	none	xgbTree	43	145	42	1041	0.853	0.692	0.315	0.229	0.506	0.751	0.253	0.503
M2	none	xgbTree	40	148	39	1044	0.853	0.691	0.300	0.213	0.506	0.766	0.256	0.495
M1	none	xgbLinear	49	139	87	996	0.822	0.619	0.302	0.261	0.360	0.727	0.241	0.461

The top-performing model for each method is documented in Table 6.4. Initially, the top models, as outlined in section 6.7.4, underwent filtration based on the 0.50 quantile for both the Brier score and AUC for each method. Subsequently, they were arranged by their

performance score, with only the top model being presented in the table. Baseline 2 did not yield satisfactory results. Among the remaining methods, XGBTree and XGBLinear models demonstrated superior performance. The proposed methods outperform the Baseline models, with a relatively even comparison among the three proposed models. Method 2 excelled in the F1 score, while Method 3 exhibited the highest balanced accuracy. However, F1 scores under 0.5 are not considered to be good.

Table 6.4: Aim 3 - Top Performance Results of Each Predictive Method for Object Localization

m.	ss.	model	TP	FP	FN	TN	acc.	b-acc	F1	prc.	rec.	AUC	Brier	prfm.
B1	none	rule-based	53	140	135	943	0.784	0.576	0.278	0.275	0.282	0.670	0.216	0.427
B2	none	dt_gini	0	188	0	1083	0.852			0.000		0.500	0.309	
B3	none	xgbTree	50	138	64	1019	0.841	0.660	0.331	0.266	0.439	0.689	0.269	0.495
M1	none	xgbTree	56	132	69	1014	0.842	0.666	0.358	0.298	0.448	0.733	0.243	0.512
M2	none	xgbLinear	65	123	96	987	0.828	0.646	0.372	0.346	0.404	0.755	0.240	0.509
M3	none	xgbTree	43	145	42	1041	0.853	0.692	0.315	0.229	0.506	0.751	0.253	0.503

(2) Object Recognition

The four primary performance metrics of all methods except Baseline 1 are illustrated in Figure 6.11. Notably, we adjusted the Brier score to "1 - Brier score" to maintain consistency across the measures, ensuring that a value closer to 1 signifies better model performance.

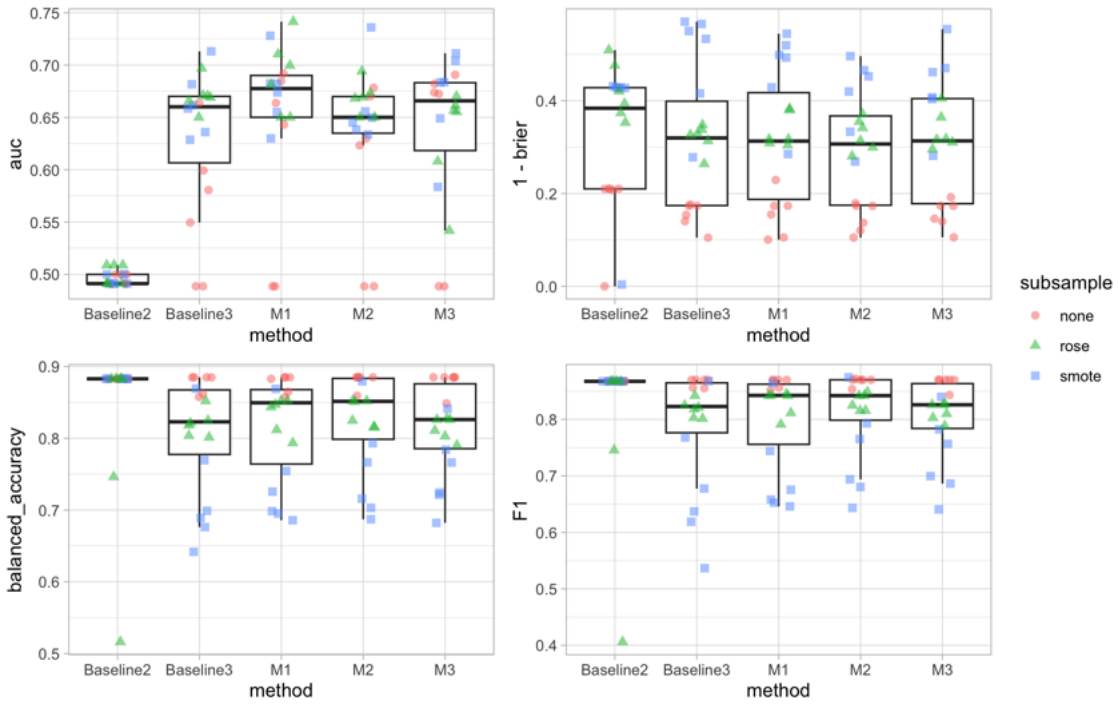


Figure 6.11: Aim 3 - Performance measures for object recognition methods

Models employing SMOTE and ROSE subsampling methods exhibit superior performance compared to non-sampled data in terms of AUC and Brier score, whereas models based on non-subsampled data excel in terms of balanced accuracy and F1 score. This validates our approach of initially filtering out models based on AUC and Brier score and subsequently comparing the model performance score, leading to the models in the final pool performing better across all dimensions. Baseline 2 continues to demonstrate the poorest performance. The remaining methods perform consistently across all four measures.

The top five models, ranked by performance score after filtering out those that exceeded the 0.85 quantile for both Brier score and AUC are presented in Table 6.5. This is the only case that a different threshold for quantile is used, for there are no valid model results that meet the requirement of exceeding 0.85 quantile for both performance measures, the threshold is relaxed for this single case. When comparing balanced accuracy, where there's a baseline of 50% from always predicting negative, the model achieving the highest

balanced accuracy is 62.8%. Regarding AUC, a range of 0.7 to 0.8 is typically considered acceptable [66]. F1 scores above 0.7 or 0.8 are often considered good in many practical scenarios [94]. Additionally, we are not displaying the best balanced accuracy or F1 score. Models that performed poorly in terms of AUC and Brier score have been filtered out, which may perform well in terms of balanced accuracy or F1 score at the cost of poor performance for the rest. It is noteworthy that all the models in the top five are based on SMOTE subsampled data.

Table 6.5: Aim 3 - Top Five Performance Results of All Predictive Methods for Object Recognition

m.	ss.	model	TP	FP	FN	TN	acc.	b-acc	F1	prc.	rec.	AUC	Brier	prfm.
M3	smote	xgbTree	98	46	17	27	0.665	0.611	0.757	0.681	0.852	0.711	0.593	0.684
M1	smote	dt_entropy	96	48	18	26	0.649	0.597	0.744	0.667	0.842	0.674	0.571	0.670
M1	smote	rForest	77	67	7	37	0.606	0.636	0.675	0.535	0.917	0.682	0.501	0.656
M3	smote	rForest	81	63	11	33	0.606	0.612	0.686	0.562	0.880	0.683	0.530	0.649
M3	smote	xgbLinear	85	59	14	30	0.612	0.598	0.700	0.590	0.859	0.684	0.538	0.649

The top-performing model for each approach is documented in Table 6.6. Initially, the leading models, as detailed in section 6.7.4, underwent filtering based on the 50% quantile concerning both the Brier score and AUC for each method. Subsequently, they were organized by their performance score, with only the top model being included in the table. Baselines 1 and 2 demonstrate inferior performance compared to the others across all dimensions. The remaining methods perform consistently. Random forest models emerge as the best performers overall. Notably, four methods achieve an F1 score exceeding 0.8, which is considered excellent.

Table 6.6: Aim 3 - Top Performance Results of Each Predictive Method for Object Recognition

m.	ss.	model	TP	FP	FN	TN	acc.	b-acc	F1	prc.	rec.	AUC	Brier	prfm.
B1	none	rule-based	40	13	104	31	0.378	0.491	0.406	0.755	0.278	0.670	0.622	0.449
B2	rose	rForest	40	104	13	31	0.378	0.492	0.406	0.278	0.755	0.509	0.524	0.449
B3	rose	rForest	117	27	24	20	0.729	0.628	0.821	0.812	0.830	0.671	0.652	0.724
M1	rose	rForest	114	30	23	21	0.718	0.622	0.811	0.792	0.832	0.700	0.617	0.717
M2	rose	xgbTree	115	29	23	21	0.723	0.627	0.816	0.799	0.833	0.673	0.645	0.721
M3	rose	rForest	113	31	22	22	0.718	0.626	0.810	0.785	0.837	0.670	0.595	0.718

6.8 Discussion

Overall, the proposed methods demonstrate consistent performance and surpass the baseline models in both phases, underscoring the importance of integrating gaze-point-based features and visual sensory-dependent features into predictive models for object-level driver SA.

Models utilizing non-subsampled data exhibit superior performance for object localization, whereas models based on subsampled examples perform better for object recognition. This difference may be attributed to the smaller size of the dataset for object recognition. Subsampling effectively augments the dataset, enhancing model performance.

It is noteworthy that Baseline 3 performs comparably well with the proposed method for object recognition, suggesting that incorporating environmental, object, and driver features is crucial for accurate object recognition prediction, which is not the case for object localization.

Chapter 7

AIM 4: PREDICT DRIVERS' SA USING REAL-WORLD DRIVING VIDEOS

This chapter covers Aim 4 of this study introduces a method for predicting SA using using real-world driving videos.

For Aim 4, predictive models of SA are built based on real-world driving videos without using eye-tracking data, which is depicted as "our model" in Figure 7.1. A thorough literature review has been done on predictive models for SA in Section 2.3. We noted that all the models used features generated from the eye-tracking data, which are denoted as "SA predictive models" in Figure 7.1. We also built one for binary and one for multi-level SA for Aim 3. However, it is not practical for now to let drivers wear an eye-tracker during driving. Eye trackers are expensive, uncomfortable to wear especially for near or far-sighted users, and require calibration thus mainly used in lab environments. Additionally, capturing and recording eye movement may cause privacy concerns. Considering that most vehicles nowadays have ADAS and dash cams, video data from the outside will be easy to obtain. What if extracting eye gaze features from these videos instead? There have been studies in the computer vision (CV) field about visual attention modeling that predicts human eye fixation from video or image data, which well fits our needs. A review of related studies is presented in subsection 6.3.1, with the DeepGazeIIE model selected as the model for gaze estimation. The visual attention matrix for each scenario was extracted and gaze points were estimated as presented in subsection 7.2. Response and predictors are introduced in subsections 7.3 and 7.4, followed by the introduction of proposed methods for both the object localization and recognition phases. Model results are exhibited in subsection 7.5 followed by a discussion on the results.

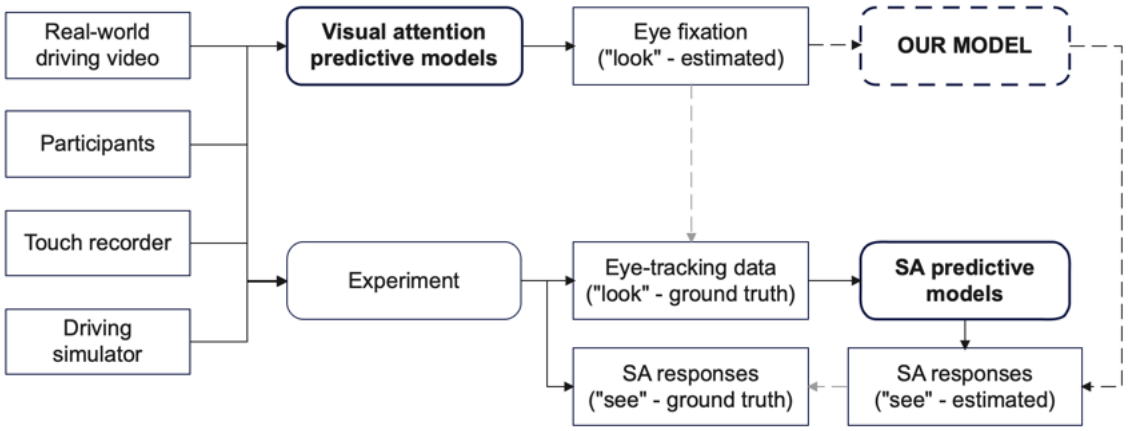


Figure 7.1: Framework of Aim 4

7.1 Driver Visual Attention Modeling Studies

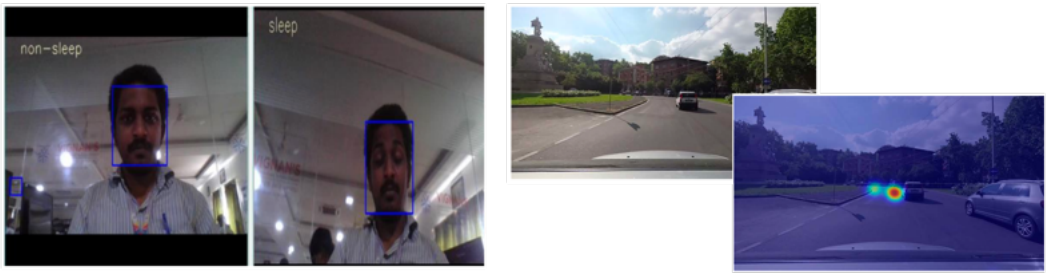
Visual attention is a field of research in image processing and computer vision closely linked to subjects such as saliency detection of objects and eye fixation [59]. Related studies can be classified into two perspectives (see Figure 7.2). First is driver-focused, which estimates whether the driver was paying attention to the driving task, such as fatigue or distraction detection, by analyzing drivers' face videos. The other is scene-focused estimating where the driver focuses using the forward roadway video. The latter is our focus.

7.1.1 Pull Strategy

The study review followed a modified Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method to identify relevant literature [71]. Two databases were searched: Web of Science and Google Scholar, selected for their coverage across science, engineering, and other disciplines related to our review topic. The search syntax and Boolean operators "OR" and "AND" were used to search in the databases:

```
(("driver") OR ("driving")) AND (("attention") OR ("perception") OR
("estimation") OR ("fixation map") OR ("fixation area") OR
("fixation region")) AND (("deep learning") OR ("neural network")) AND
```

```
(("eye tracker") OR ("eye movement") OR ("eye tracking") OR ("gaze"))
```



(a) Driver-focused example: drowsiness detection

(b) Scene-focused: top shows the scenario frame, bottom overlays the eye fixation heatmap.

Figure 7.2: Examples of related studies from two perspectives (images from [16, 23])

7.1.2 Review on Existing Methods

After screening the abstracts, we filtered out three papers out of 33 from Web of Science and 16 out of 120 from Google Scholar. As the three papers from Web of Science were duplicates of those from Google Scholar, 16 papers were left to be reviewed in full length. These scene-focused studies use scene images or videos captured by dash cams as inputs; capture drivers' eye fixation data as ground truth; and apply convolutional neural network (CNN) with different structures to estimate drivers' eye fixation map. This study flow together with examples of inputs, intermediate products, and outputs are depicted in Figure 7.3.

Palazzi et al.'s DR(eye)VE project [77] is widely recognized as the leading study in the field. They created and publicly released the first dataset addressing drivers' focus of attention prediction and named it with their project name DR(eye)VE. The dataset comprises 74 sequences containing a total of 555,000 frames collected from a naturalistic driving (ND) study. Each sequence is 5 minutes in duration and features eight drivers, aged between 20 and 40, including seven men and one woman. Each driver drove for over two months. Videos were recorded in various contexts, including downtown areas, countryside roads, and highways, as well as in different traffic conditions ranging from low to high levels

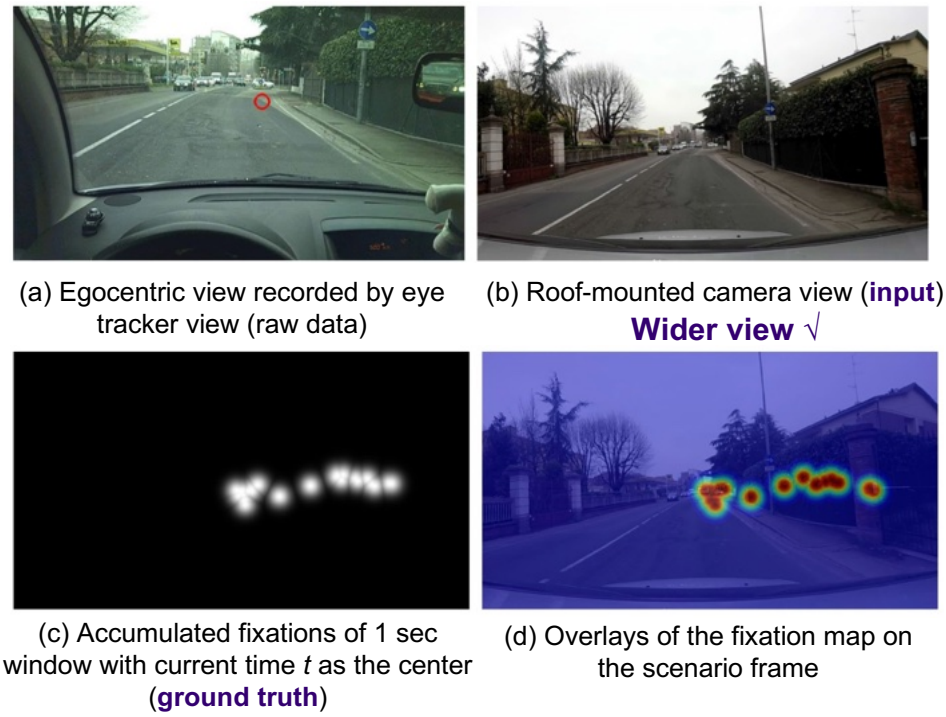


Figure 7.3: Flow if scene focused study recreated with images from [77]

of congestion; also under different weather conditions such as sunny, rainy, and cloudy, and at various times of the day including both daytime and nighttime.

A multi-branch model composed of three identical branches depicted in Figure 7.4 was used in the DR(eye)VE project. Each branch exploits complementary information from a different domain and contributes to the final prediction accordingly. In detail, the first branch works in the RGB domain and processes raw visual data about the scene, denoted as X_{RGB}. The second branch focuses on motion through the optical flow, denoted as X_{flow}. The last branch takes as input semantic segmentation probability maps, denoted as X_{seg}. The model proposed outperforms several baselines and competitors. Results from the study indicate that there are common attention patterns among drivers, which can be partially replicated.

Studies in this field can be grouped based on five different dimensions:

- *Data source*: 1) directly use data from existing datasets, especially the DR(eye)VE

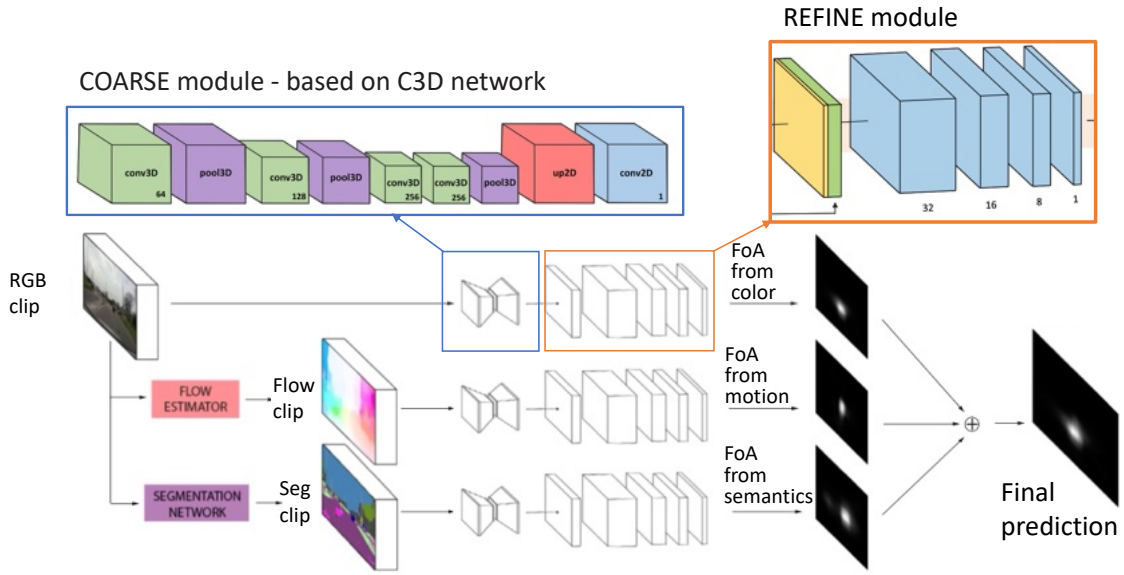


Figure 7.4: The architecture of a multi-branch deep network for the focus of attention (FoA) proposed by DR(eye)VE project [77] recreated based on images from the paper

dataset and the Berkeley DeepDrive Attention (BDD-A) dataset[1]; 2) collect data using the scenarios from existing studies with the video-mimicking method; 3) collect data fully on their own including ND data collected from real-world driving, video-mimicking, CG-based, or CG-mimicking collected in lab environments.

- *Input*: 1) raw scenario image or video (i.e., image stream); 2) processed scenario image or video including optical flow, semantic segmentation, and saliency; 3) others such as gaze direction, physiological data, and demographic data.
- *Output*: 1) estimated eye fixation map for each pixel; 2) estimated fixation region/zone for each region/zone (uniformly distributed in each region or zone).
- *Model Architecture*: 1) single-stream CNN with single channel input aiming for faster running speed; 2) multi-stream/modal/branch CNN with multiple channels of inputs aiming for higher accuracy; 3) skip layer CNN that tries to find a balance between speed and accuracy; 4) generative adversarial network (GAN).

[1] https://bdd-data.berkeley.edu/

- *Model comparison*: Assess the performance of the model relative to competitors under two conditions: 1) exclusively in the driving context, and 2) in a generic context. In a driving context, drivers typically exhibit a central fixation bias towards the forward roadway, which distinguishes it from other contexts. Although several attention prediction models exist without specifying a context, their effectiveness in a driving context may be limited.

Detailed notes for the papers are summarized in a table in Table 7.1.

7.1.3 Issues Regarding Applying Existing Studies

Upon discovering the pre-trained DReyeVE [2] model and noting its exceptional performance, which aligned perfectly with our requirements, we opted to utilize it. However, we were disappointed to find that although the model was provided as a *.h5* file, it was implemented using *theano*, which is no longer actively maintained and has been succeeded by *tensorflow*. Configuring the Python environment was challenging as well due to significant advancements in coding languages and their associated packages in recent years.

As a result, we conducted further research to identify the latest models that suited our objectives. While we identified some recent studies, they were not specifically tailored for driving applications. These studies highlighted transfer learning as a pivotal factor in improving spatial saliency prediction since 2014, although progress has stagnated over the past 3-5 years [58, 61].

Ultimately, we discovered the DeepGazeIIE [3] model, which offered straightforward implementation and demonstrated superior performance on both datasets: the MIT 3000 that has 300 natural indoor and outdoor scenes with human eye tracking data; and CAT 2000 which contains 4000 images from 20 different categories with human eye tracking data. This model outperformed over 40 others listed on the "MIT/Tuebingen Saliency Benchmark" website [4], which serves as an up-to-date online resource for saliency model performances

[2] https://github.com/ndrplz/dreyeve

[3] https://github.com/matthias-k/DeepGaze

[4] https://saliency.tuebingen.ai/results.html

and datasets. The contributors maintain a continually updated comparison page, serving as a valuable resource for documenting and advancing progress in the field of saliency prediction, i.e., eye gaze estimation.

Table 7.1: Reviewed Studies on the Prediction of Driver Visual Attention

Study∗	**Year**	**Data Source**	**Input**	**Output**	**Model**	**CNN Architecture**	**Model Comparison**
Deng [23]	2019	Self-collected ND data, tested on DR(eye)VE	scenario video	fixation map	CDNN	skip layer	generic, driving
Deng [22]	2021	self-collected ND data	scenario video	fixation map	DSTANet	skip layer	generic, driving
Gou [39]	2022	BDD-A dataset	scenario video	fixation map	ACT-Net	skip layer	generic, driving
Hong [82]	2021	Self-collected ND data, tested on DR(eye)VE	scenario video, gaze region	fixation region	DFaeP	single stream	generic
Hou [46]	2022	self-collected, CG-based	simulator scenario video, gaze region	fixation region	VGG16	single stream	-
Hu [47]	2021	self-collected, video mimicking with data from DR(eye)VE data in virtual reality (VR) env.	scenario image - processed: saliency inference, dense optical flow, semantic segmentation + gaze direction	fixation map	Multi-resolution Neural Network, 3D-CNN	multi-stream	generic, driving
Huang [48]	2022	self-collected, video mimicking with data from DADA-2000	scenario image stream	fixation map	DDD (a 3D-CNN)	single stream	driving
Kang [54]	2020	DR(eye)VE	scenario video: multi-scale or multiple resolutions	fixation map	FCN	skip layer	generic, driving
Lteef et al.,	2020	BDDA, DR(eye)VE	scenario image	fixation map	GAN	-	generic, driving
Makrigiorgos [65]	2019	self-collected, CG-based in VR env.	scenario image: raw images, dense optical flow images	fixation map	based on DR(eye)VE	multi-stream	generic, driving
Ning [75]	2022	DR(eye)VE	scenario image: raw images, dense optical flow images	fixation map	Y-FCNN (fuse two VGG16)	multi-stream	driving
Palazzi [77]	2018	self-collected ND data and built DR(eye)VE dataset	scenario video: raw, dense optical flow, semantic segmentation	fixation map	DR(eye)VE	multi-stream	driving
Qiu [84]	2020	self-collected, CG-based	scenario video (semantic segmentation) + fixation map + CAN-Bus signals (steering wheel, speed) + Physiological signals (heart rate, GSR: galvanic skin response) + navigation info + transparency info (how much operation data were provided) + weather info	takeover decision	3D-CNN	multi-stream	-
Shirpour [92]	2021	self-collected ND data	scenario image + gaze direction	fixation map	based on VGG16	skip layer + multi-stream	generic

∗ Only the first author is listed considering the limited space.

7.2 Visual Attention Data Processing

7.2.1 Image Blending

Optical flow or optic flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and a scene [6]. It is very typical for drivers to observe optical flows rather than static scenes while driving. Optical flow requires heavy computation and image blending can be a cheap and practical substitute for this study.

Image blending, also known as image fusion or image composition, refers to the process of combining two or more images into a single composite image [11]. This technique is commonly used in computer vision, remote sensing, medical imaging, and various other fields to enhance or extract useful information from multiple sources.

Denote the number of frames pre-pause as N_f, then the frame chain we have can be written as: $img_{-N_f}, img_{-N_f+1}, \ldots, img_{-N_f+(N_f-1)}, img_{-N_f+(N_f)}$. Denote the weight between the blended image of the first $k-1$ frames and the kth frame as α, then the kth blended image could be written as:

$$\begin{aligned}\text{Blended Image}_k =& \alpha \times (\text{Blended Image}_{k-1}) + (1-\alpha) \times img_{-N_f+k} \\ =& \alpha^k \times img_{-N_f} + (1-\alpha)\alpha^{k-1} \times img_{-N_f+1} + (1-\alpha)\alpha^{k-2} \times img_{-N_f+2} \\ & \cdots + (1-\alpha)\alpha \times img_{-N_f+(k-1)} + (1-\alpha) \times img_{-N_f+k}\end{aligned} \tag{7.1}$$

The final blended images of N_f pre-pause frames can be written as:

$$\begin{aligned}\text{Blended Image}_{N_f} =& \alpha \times (\text{Blended Image}_{N_f-1}) + (1-\alpha) \times img_0 \\ =& \alpha^{N_f} \times img_{-N_f} + (1-\alpha)\alpha^{N_f-1} \times img_{-N_f+1} + (1-\alpha)\alpha^{N_f-2} \times img_{-N_f+2} \\ & \cdots + (1-\alpha)\alpha \times img_{-1} + (1-\alpha) \times img_0\end{aligned} \tag{7.2}$$

Where img_0 is the scene image at pause. The weight of image img_{-k} will be $(1-\alpha)\alpha^{N_f-k}$. The final blended images of N_f equals 30, 60, and 90 have been created with α equals 0.5 and 0.6. According to preliminary observations, the output blended images of different N_f look almost the same. This is because the weight of the images from img_{-90} to img_{-30} are far less than the images from img_{-30} to img_0 (see Table 7.2). Examples of final blended

images of N_f equals 30, 60, and 90 with α equals 0.5 for scenario no.16 are shown in Figure 7.5. The distinction between the frame at pause and the blended images is readily apparent, whereas the blended images created from varying numbers of images appear identical. This observation was later confirmed through an examination of the visual attention (VA) metrics calculated for each blended image, revealing a correlation coefficient of one. Therefore, we only created and used images blended from 30 pre-pause frames in addition to the frame at pause in the next step – VA estimation.

Table 7.2: The weights of images img_{-30}, img_{-60} and img_{-90}

	α^{N_f}		
α	$N_f = 30$	$N_f = 60$	$N_f = 90$
0.1	1.00e-30	1.00e-60	1.00e-90
0.2	1.07e-21	1.15e-42	1.24e-63
0.3	2.06e-16	4.24e-32	8.73e-48
0.4	1.15e-12	1.33e-24	1.53e-36
0.5	9.31e-10	8.67e-19	8.08e-28
0.6	2.21e-7	4.89e-14	1.08e-20
0.7	2.25e-5	5.08e-10	1.15e-14
0.8	1.24e-3	1.53e-6	1.90e-9
0.9	4.24e-2	1.80e-3	7.62e-5

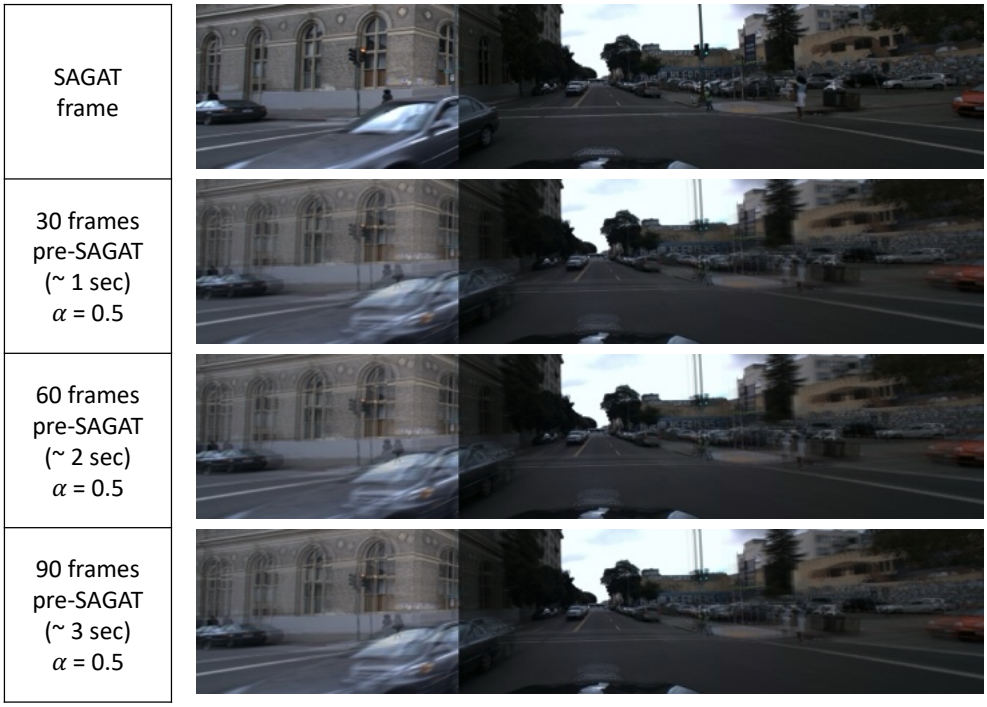


Figure 7.5: Example blended images for different numbers of frames pre-pause with equal weights for the previous blended image and the next image (scenario no.16)

An example of blended images from 30 pre-pause frames with different choices of image weights for scenario no.20 is displayed in Figure 7.6. We can see the greater the α value is, the more "blur" the peripheral views are, indicating that we rely more on the past. It is noteworthy that the image at pause can be considered as a special case with $\alpha = 0$, indicating no weights allocated to the pre-pause frames.

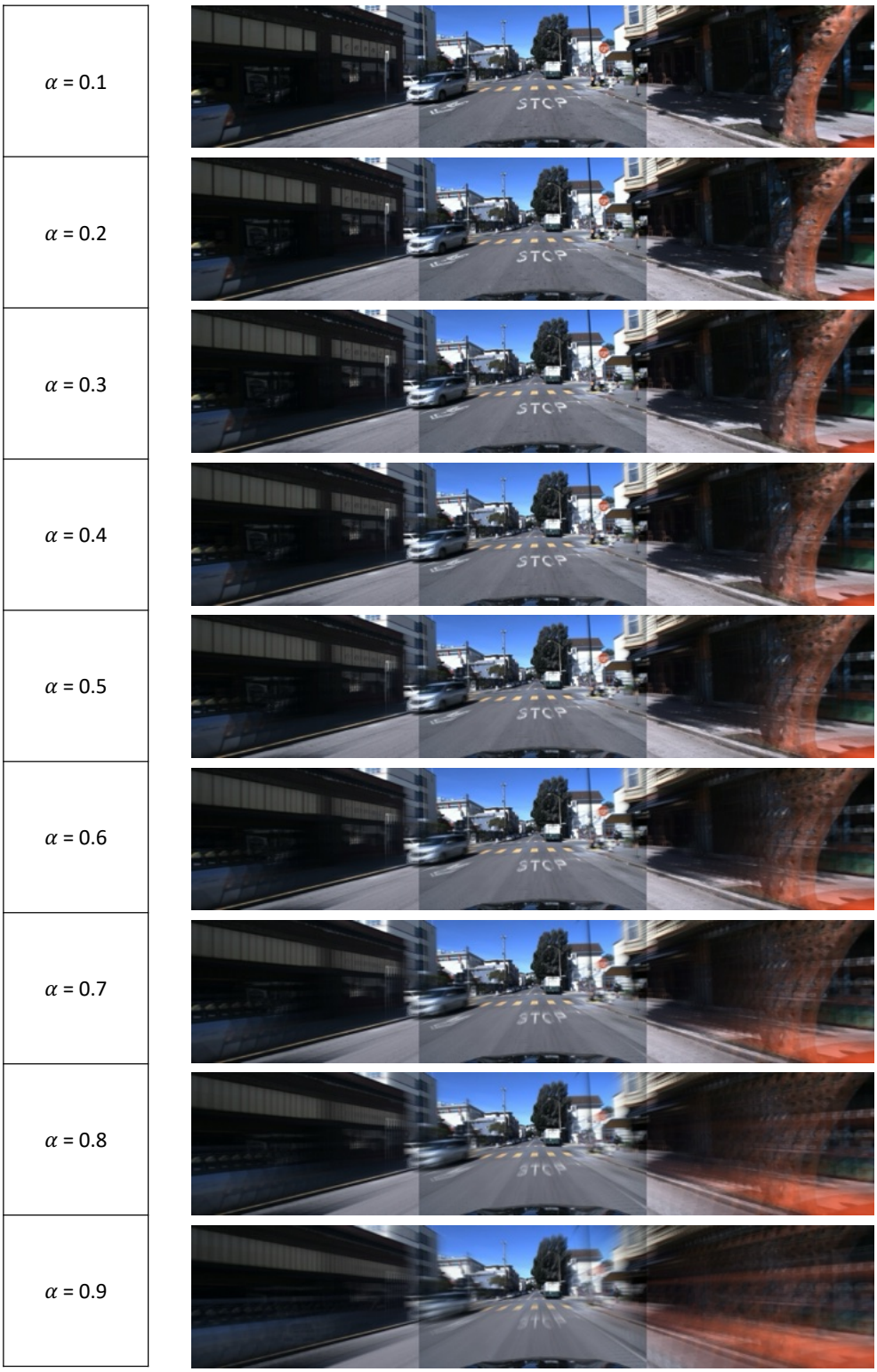


Figure 7.6: Blended image for 30 frames pre-pause with different weights for images (scenario no.20)

7.2.2 Visual Attention Estimation

The visual attention of each scenario was estimated using the DeepGazeIIE[5] model. The output is a matrix with 600 rows and 2880 columns, where each cell contains a value between zero and one. The value indicates the probability of being looked at by the driver. Overall, this matrix represents a two-dimensional probability distribution. The size of the matrix is half the size of the original image (width = 5760 px, height = 1200 px) and was resized because that's the highest resolution the laptop's GPU could handle. Figure 7.7 shows two 3D surface plot examples for the visual attention matrix.

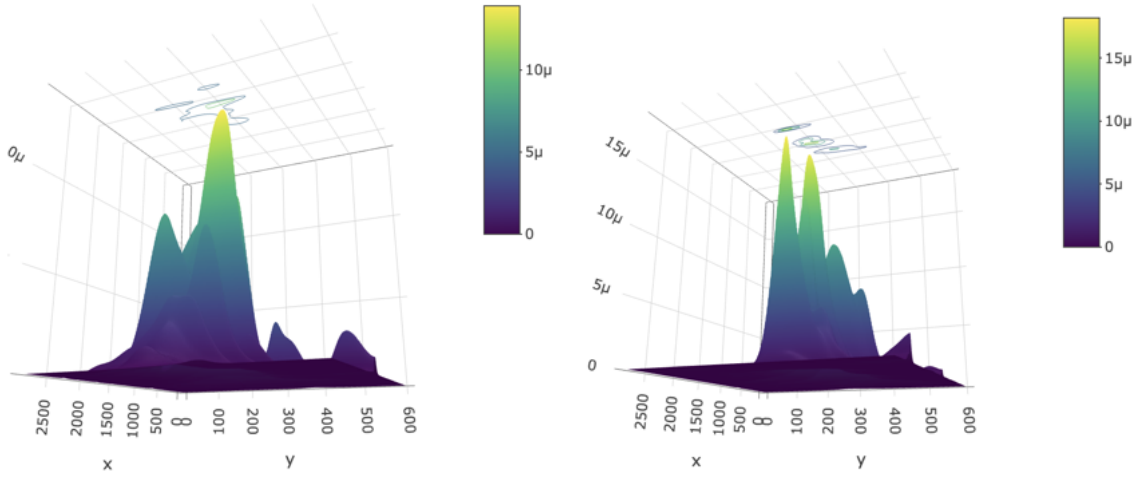


Figure 7.7: 3D surface with contour plot examples for VA matrix. Scenario no.16 on the left, and scenario no.20 on the right.

[5]https://github.com/matthias-k/DeepGaze

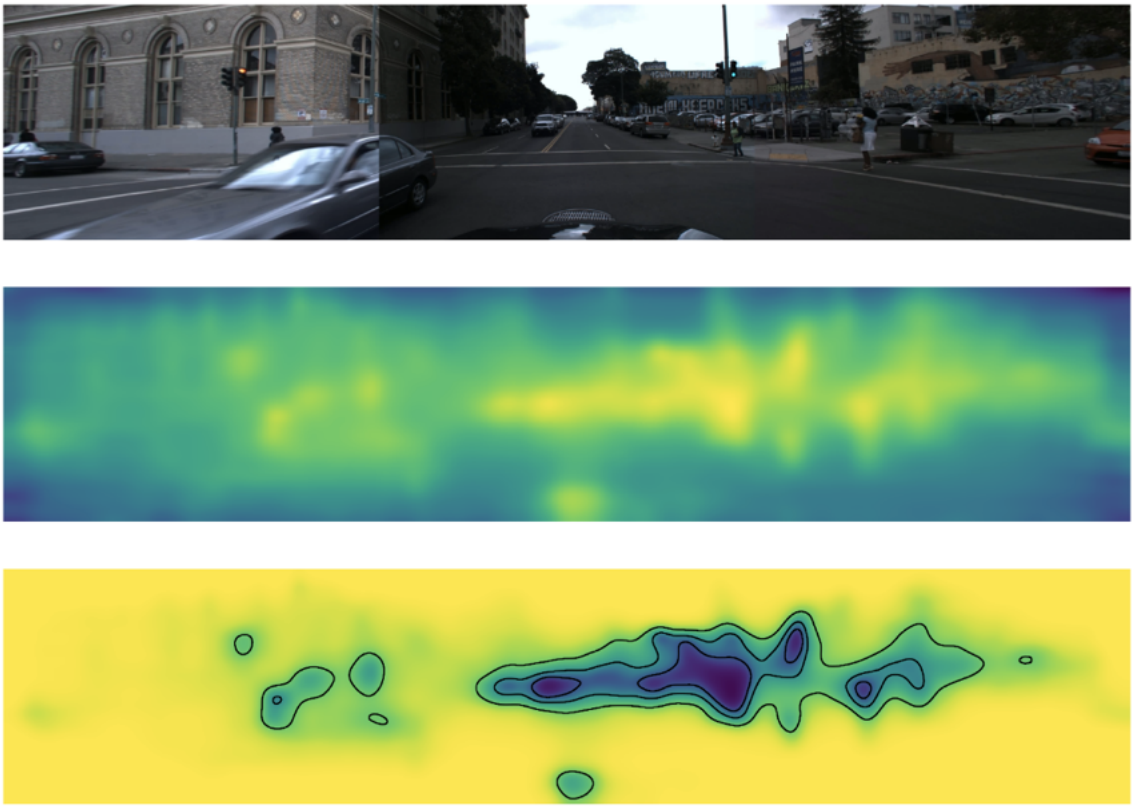


Figure 7.8: Example of DeepGazeIIE model output for scenario no.16: input image, density map, and density map with contours.

7.3 Response

The response variables are consistent with Aim 3, and two-phase predictive models will be applied. SA regarding object localization will be predicted first, followed by the SA regarding object recognition.

Object-wise SA will be the focus of the proposed work considering our research goal to selectively provide assistance.

7.3.1 SA regarding Object Localization

The first-phase models predict the situation awareness regarding object localization by determining whether the object's location has been accurately identified. The SA response

indicating object localization is a binary variable for the first model as depicted in Figure 6.1.

- *Hit* - the location of the object has been correctly identified. If any touch has been detected within the object's bounding box area or its trajectory, the object's location.
- *Miss* - the object was not detected by the participant. No touch located within the object's bounding box or its trajectory has been detected.

7.3.2 SA regarding Object Recognition

The second-phase models predict the situation awareness regarding object recognition by determining whether the object's type has been accurately identified. The SA response indicating object recognition is a binary variable for the second model:

- *Correct Type* - both the object location and its type have been correctly identified.
- *Wrong Type* - the location of the object was accurately recognized, but its type was misidentified.

7.4 Predictors

7.4.1 Visual Attention Metrics

(a) Object-level visual attention density

The object-level VA density was calculated by summing and then dividing the values extracted from the VA matrix within the object's bounding box by the size of the bounding box, see equation 7.3. Initially, the sum served as the metric. However, it was noted that this metric, when computed from the image at pause, showed a correlation coefficient of 0.62 with object size, while when computed from the blended image of 30 frames pre-pause, the correlation coefficient was 0.64. As a result, density was computed and employed as the predictor, resulting in correlation coefficients of 0.22 and 0.24 with object size, respectively.

$$VA_{Density} = \frac{\sum_{(x,y)\in bbox_i} VA_{xy}}{OO_{size}} \tag{7.3}$$

where VA_{xy} represents a cell in the xth row and yth column of the VA matrix; $bbox_i$ refers to the bounding box of object i and OO_{size} refers to the size of object i.

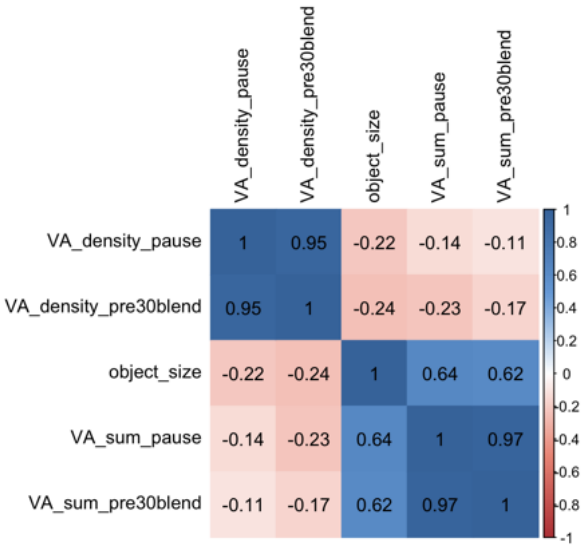


Figure 7.9: Correlation between object's accumulated visual attention probabilities (predictors starts with `VA_sum`), object's VA density (predictors starts with `VA_density`), and object size. `pause` indicates the VA matrix was computed based on the frame at pause; `pre30blend` indicates the matrix was computed based on the 30 frames pre-pause in addition to the frame at pause.

(b) Estimated gaze-point-based features

It is essential to estimate the eye gaze points first in order to calculate the gaze-point-based features. Figure 7.10 outlines the steps for gaze point estimation. Initially, the DeepGazeIIE model is utilized to predict the probability distribution of eye gaze for each input image, which can either be an image at pause or a blended image of multiple frames pre-pause. Subsequently, continuous regions are identified, and the centroid coordinates of each region are captured as the estimated gaze points. The first two steps have been thoroughly discussed in section 7.2. Further elaboration on the remaining two steps is provided in the subsequent paragraphs.

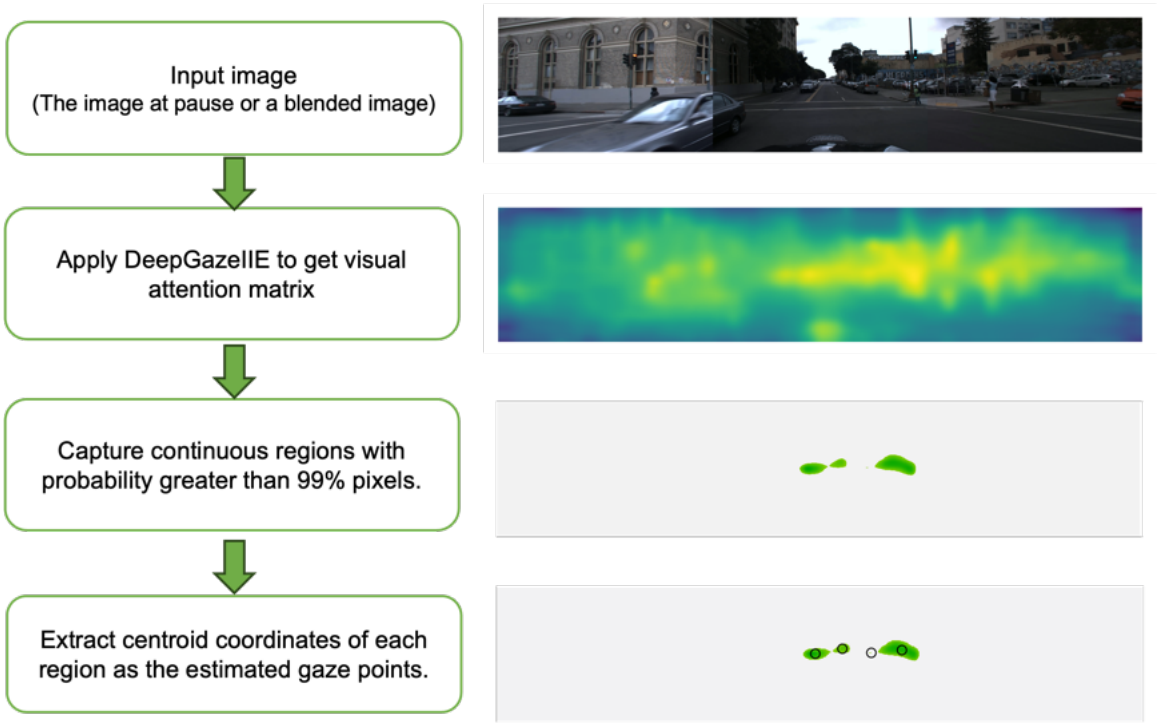
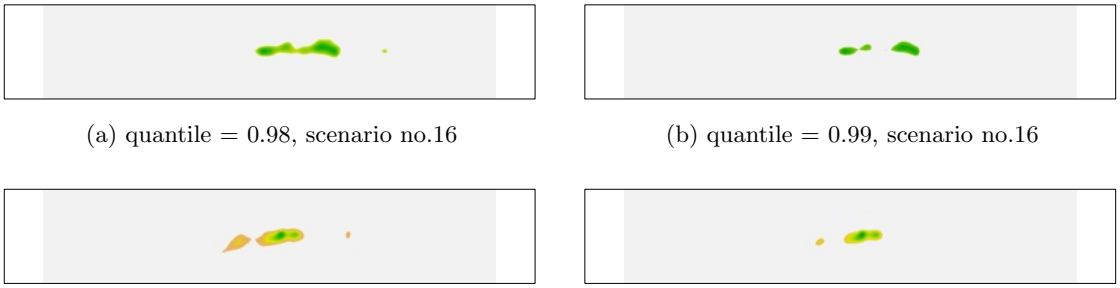


Figure 7.10: Flow chart for gaze point estimation procedures

b.1. Capture Continuous Regions

The peaks in a VA matrix were regarded as the estimated gaze points, identified by estimated probabilities of being looked at by the driver surpassing the 0.99 quantile of all pixels. This pool comprises $600 \times 2880 \times 1\% = 17280$ pixels. Considering the memory capabilities of drivers, it is reasonable to have multiple estimated gaze points, denoted by multiple peaks in the VA matrix, within a scene. This ensures an accurate representation of areas they have observed, rather than merely glanced at.

Various quantiles ranging from 0.95 to 0.99 were examined to empirically determine the optimal threshold. As demonstrated in Figure 7.11, a threshold of 0.99 yields more precise delineations of local peak regions compared to 0.98. Smaller quantiles yield even larger regions compared to 0.98 and examples are not presented in this dissertation.



(a) quantile = 0.98, scenario no.16

(b) quantile = 0.99, scenario no.16

(c) quantile = 0.98, scenario no.20

(d) quantile = 0.99, scenario no.20

Figure 7.11: Heatmaps illustrating points exceeding respective quantiles.

b.2. Extract Centroid Coordinates of Continuous Regions

Upon the continuous regions being extracted from the pixels filtered with the threshold, the coordinate of the centroid of each region R_i is computed with the following equation.

$$(x,\ y)_{R_i}^{centroid} = \frac{\sum_{(x,\ y)\in R_i}(x,\ y)}{N_{(x,\ y)\in R_i}} \tag{7.4}$$

where x and y refer to the coordinates of points within region R_i. The output centroids are the estimated gaze points used to compute the gaze-point-based and visual sensory-dependent features.

b.3. Compute Gaze-Point-Based Features

Gaze-point-based features use the positional relationship between target objects and participants' eye fixation center coordinates. Distance between the eye fixation and its closest point to the bounding box has been used to compute the following two types of object-gaze distance:

- $\widehat{FD}_{min}^{\alpha}$ - minimum gaze distance to an object within the analysis time window T, computed using estimated gaze points, measured in degree.
- $\widehat{FD}_{avg}^{\alpha}$ - average gaze distance of an object during the analysis time window T, computed using estimated gaze points, measured in degree.

where α represents the weight we used for the blended image.

(c) Select feature sets extracted from images blended with different weights α

Correlation coefficients between features extracted from images blended with different weights were examined, including both the VA density as shown in Figure 7.12, and gaze-point-based features as shown in Figure 7.13.

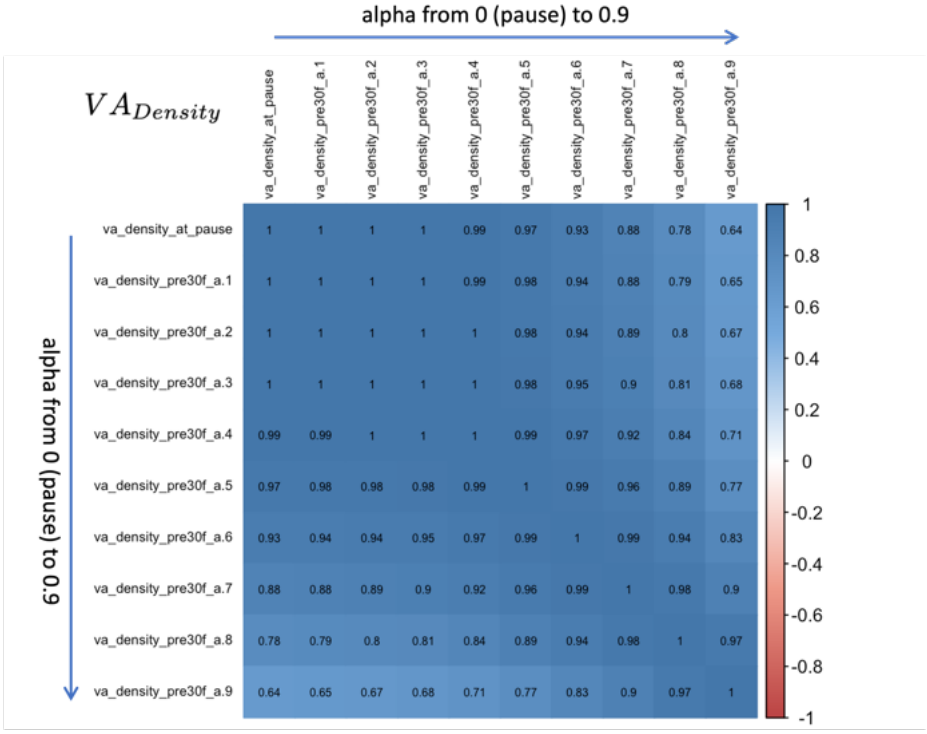


Figure 7.12: Correlation coefficients between VA density extracted from images blended with different weights

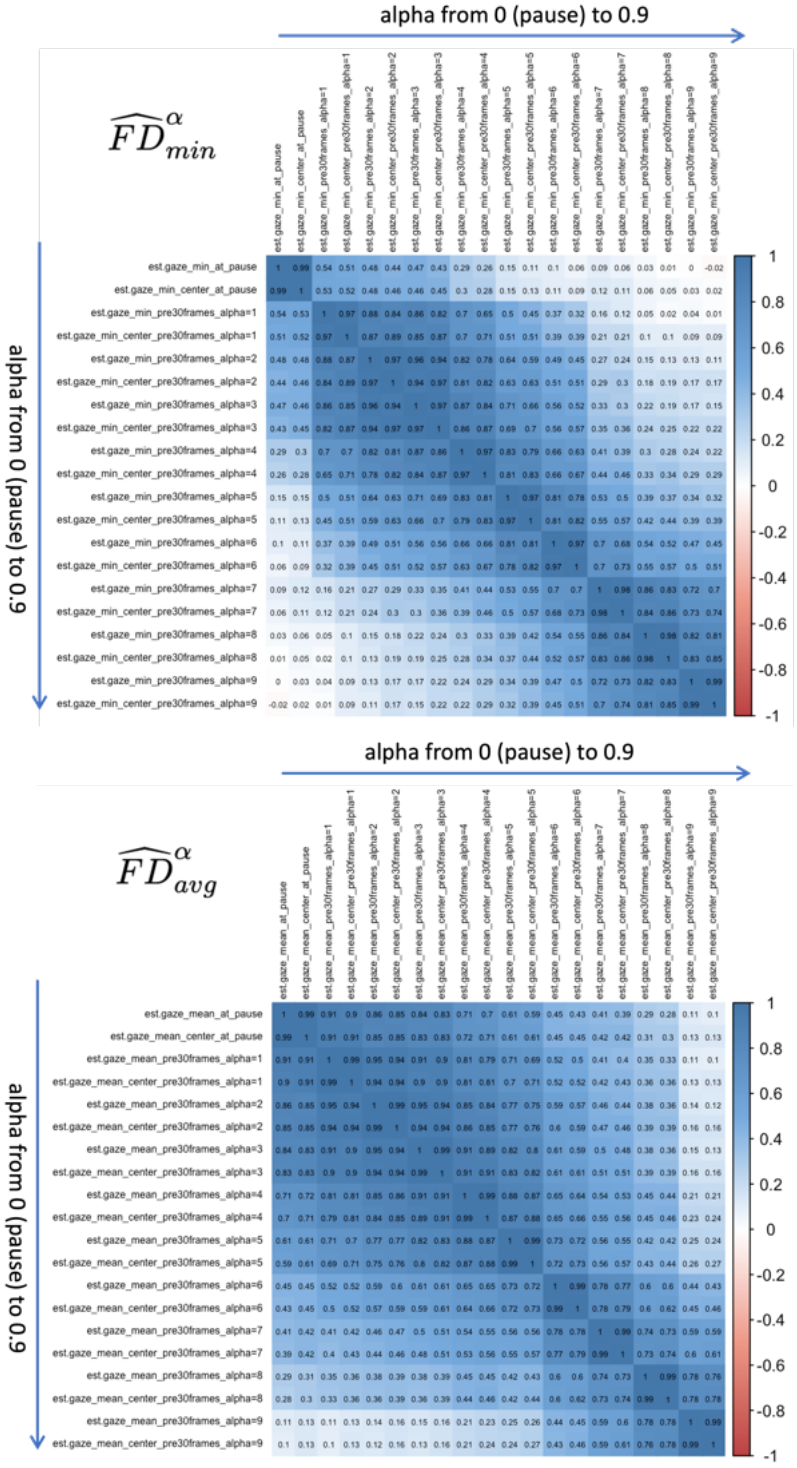


Figure 7.13: Correlation coefficients between gaze-point-based features extracted from images blended with different weights

As explained in section 7.2, a larger α assigns greater weights to the pre-pause frames that are farther from the pause, i.e., allocating more weights to the trajectory over the static scene. To explore the influence of different image blending weights while also optimizing training time, models were not trained for every weight. Given that each α value entails training: 6 proposed methods $\times$ 6 models $\times$ 3 subsampling options = 108 models, weights $\alpha = 0, 0.3, 0.6, 0.9$ were chosen for training based on correlation results presented in Figures 7.12 and 7.13. This approach allowed for a comprehensive investigation into the impact of weights on the trajectory toward SA prediction while mitigating the training time required.

Additionally, the case incorporating VA-extracted features of all weights was considered, including α from 0 to 0.9, with PCA utilized to address multicollinearity issues among predictors.

7.4.2 Environmental, Object and Demographic Features

The environmental, object, and demographic features used for Aim 4 are identical to those used for Aim 3 which have been described in section 6.3.4.

7.5 Data Split

The data split for Aim 4 is identical to Aim 3, which has been elaborated in section 6.4. Out of the 75 scenarios, 15 (20%) have been randomly assigned to the test set, while the remaining 60 (80%) are assigned to the training set. This allocation results in a training set consisting of 4921 rows and a test set of 1271 rows for object localization, and 1127 rows for training and 188 rows for testing for object recognition.

7.6 Subsampling

The subsampling for Aim 4 is identical to Aim 3, which has been elaborated in section 6.5. Three subsampling types will be considered: none, SMOTE, and ROSE.

7.7 Method Evaluation

The same predictor set will be utilized for both phases of the two-phase method set. In the first phase, the model predicts SA concerning object localization, while in the second

phase, it predicts SA regarding object recognition. A summary of the feature sets for all methods is provided in Table 7.3, with detailed descriptions available in the subsequent subsections. The numbers within the cells represent the number of predictors employed in each respective method. Additionally, numbers within brackets alongside "PCA" indicate the application of PCA, with the enclosed number denoting the count of PCA-transformed predictors utilized in the method.

The estimated gazes of images at pause ($\alpha = 0$), and blended images of 30 pre-pause frames with weights $\alpha = 0.3$, $\alpha = 0.6$, $\alpha = 0.9$ and $\alpha = all$ were trained for each method.

Table 7.3: Feature Sets of Each Method for Aim 4

Method (alpha_i = 0, 0.3, 0.6, 0.9)	Model	Est. gaze in bbox	Est. gaze point-based features	Visual attention score	Environmental features	Object features	Driver features	Predictors count
M4ai	rule-based	1						1
M4all_alpha	rule-based	1						1
M5ai	learning	1						1
M5all_alpha	learning	1						1
M6ai	learning	1			5	2	3	11
M6all_alpha	learning	1			5	2	3	11
M7ai	learning			1	5	2	3	11
M7all_alpha	learning			1 (PCA)	5	2	3	11
M8ai	learning		2		5	2	3	12
M8all_alpha	learning		7 (PCA)		5	2	3	12
M9ai	learning		2	1	5	2	3	13
M9all_alpha	learning		7 (PCA)	1 (PCA)	5	2	3	18

7.7.1 Proposed Methods

The methods outlined in this chapter adhere to the naming conventions in Aim 3, extending from Method 4.

Method 4: Estimated gaze enclosure rule-based

The underlying assumption posits that "looking implies seeing," suggesting that the direction of a driver's estimated eye gaze directly signifies whether they have perceived the object. This methodology hinges solely on whether the gaze point falls within the bounding box to forecast the driver's situational awareness. The predictor "est.gaze enclosure" serves as the sole predictor for predicting driver situational awareness using a rule-based model. The estimated gazes from images at pause ($\alpha = 0$) and blended images of 30 pre-pause frames with weights $\alpha = 0.3$, $\alpha = 0.6$, $\alpha = 0.9$, and $\alpha =$ all were utilized for each rule-based model. In the case of $\alpha =$ all, any estimated gaze computed with various blending weights enclosed in the bounding box was considered as an estimated gaze enclosure.

Method 5: Estimated gaze enclosure learning-based

The fundamental assumption aligns with Method 4 that the direction of a driver's estimated gaze directly indicates whether they have perceived the object. However, in this approach, learning-based models are employed. These models are consistent with those utilized in other methods, and the binary classification models are detailed in Section 6.7.3.

Method 6: Estimated gaze enclosure + environmental, object and driver features

This method echoes the assumptions of the previous two. The environmental, object, and driver features investigated in Aim 2 were added, and learning-based models were employed.

Method 7: VA density + environmental, object and driver features

In the method, the binary predictor "est.gaze enclosure" is replaced by estimated VA density features. The estimated gazes from images at pause ($\alpha = 0$) and blended images of 30 pre-pause frames with weights $\alpha = 0.3$, $\alpha = 0.6$, $\alpha = 0.9$, and $\alpha =$ all were utilized.

For $\alpha =$ all, given the correlation between predictors illustrated in Figure 7.12, PCA was utilized to address this issue of multicollinearity. The first principal component was chosen, accounting for 91.4% of the variances among the 10 predictors, passing the 90% threshold pre-set for this study. These predictors consist of the ten α values. The scree plot depicted

in Figure 7.14 illustrates the percentage of variance explained by each principal component, while the biplot on the right demonstrates the influence of each predictor on the first two principal components.

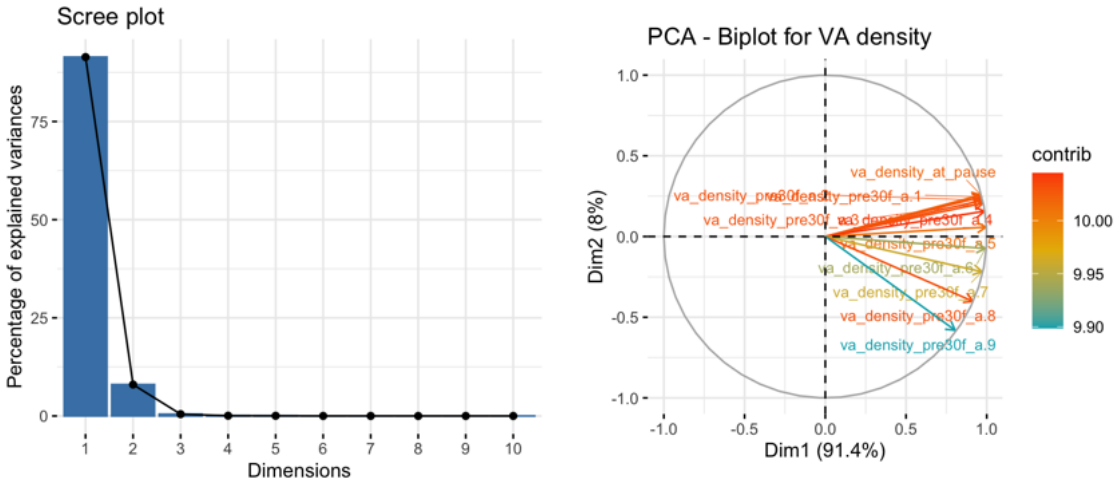


Figure 7.14: Scree plot (left) and biplot (right) of PCA results for VA density features

Method 8: Estimated gaze-point-based features + environmental, object and driver features

In the method, the binary predictor "est.gaze enclosure" is replaced by estimated gaze-point-based features. A notable correlation has been observed between all pairs of distance-related gaze-point-based features, regardless of whether they were computed using the object's center point or the nearest point on the bounding box's perimeter. Therefore, only the two features, FD_{min} and FD_{avg}, regarding the nearest point to the bounding box were incorporated.

For $\alpha =$ all, given the correlation between predictors illustrated in Figure 7.13, PCA was utilized to address this issue of multicollinearity. The first seven principal components were chosen, accounting for 91.3% of the variances among the 40 predictors, passing the 90% threshold pre-set for this study. These predictors consist of the ten α values multiplied by two estimated gaze-point-based predictions per distance, namely, distance from the center point and distance from the nearest point on the bounding box. The scree plot depicted in

Figure 7.15 illustrates the percentage of variance explained by each principal component, while the biplot on the right demonstrates the influence of each predictor on the first two principal components.

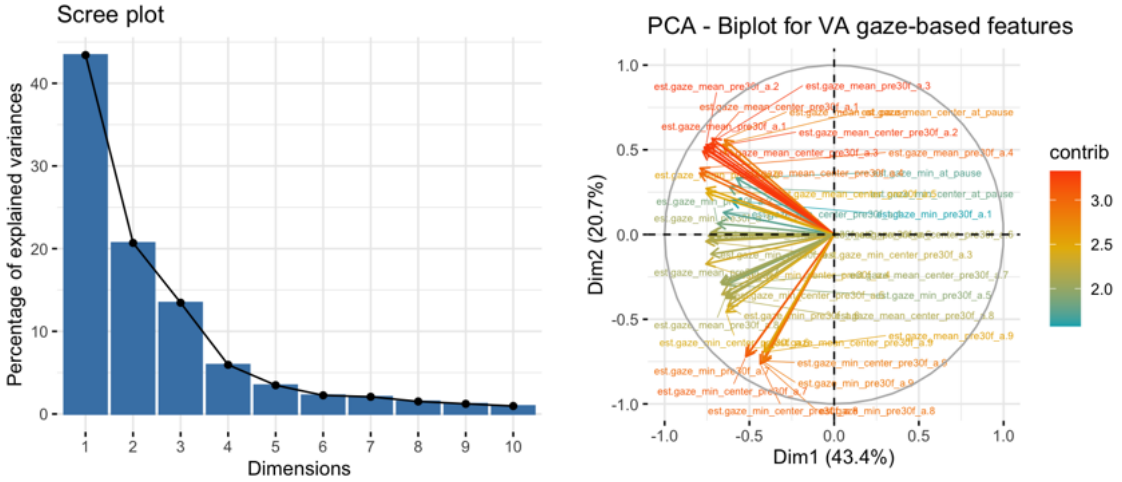


Figure 7.15: Scree plot (left) and biplot (right) of PCA results for estimated gaze-point-based features

Method 9: Estimated gaze point-based features + VA density + environmental, object and driver features

The proposed method incorporates both the estimated gaze-point-based features and the VA density features. No obvious correlation was detected between the two sets of features. Therefore, for $\alpha = 0, 0.3, 0.6, 0.9$, two features, FD_{min} and FD_{avg}, regarding the nearest point to the bounding box, were incorporated. For $\alpha = all$, PCA transformed predictors generated for Methods 7 and 8 were employed.

7.7.2 Binary Classification Models

Identical to Aim 3, as described in section 6.7.3, binary classifiers were employed for both phases of each method. The predictor sets will be the same for two phases of one method, but the binary classifiers can be different. Six learning-based classic binary classifiers have been

employed: logistic regression, decision tree (entropy, Gini), random forest, and XGBoost linear and XGBoost tree. R package *caret* (Version 6.0-94) was used for training the models. All numerical predictors have been centered and scaled pre-training. A five-fold cross-validation was employed for all training processes.

7.7.3 Model Performance Measures

Identical to Aim 3, as described in section 6.7.4. Four metrics, AUC, Brier score, balanced accuracy, and F1 were used for model comparison. Models that performed above the 0.85 quantile in terms of both the Brier score and AUC were filtered out. Subsequently, we arranged them by a performance score calculated as the average of the F1 score and balanced accuracy.

In comparing the best model for each method, models were filtered based on the 50% quantile concerning the Brier score and AUC. Since model results were grouped by method, with each group containing 90 models (except M4 which was trained with five rule-based models), significantly fewer than the total of 450 models, the quantiles were adjusted to include more models in the final pool for performance score comparison. Models were then sorted by the self-defined performance score computed as the average of the F1 score and balanced accuracy.

7.7.4 Results

Altogether, five rule-based models were executed for Method 4. For the remaining methods, Methods 5 to 9, a total of six binary classification models were trained and tested using non-subsampled data, SMOTE subsampled examples, and ROSE subsampled examples, with five α values, resulting in a combination of five times six times three times five, totaling 450 + 5 = 455 models for each phase of a method. Detailed performance results for each model can be found in Appendix A.2.

(1) Object Localization

The four primary performance metrics of all methods are illustrated in Figure 7.16. Notably, the Brier score has been adjusted to "1 - Brier score" to maintain consistency across the measures, ensuring that a value closer to 1 signifies better model performance. Models using non-subsampled data exhibit superior performance in terms of the Brier score. No significant distinction was noted among various subsampling methods. Method 5 displays the poorest performance in terms of AUC, while Method 4 excels in the Brier score.

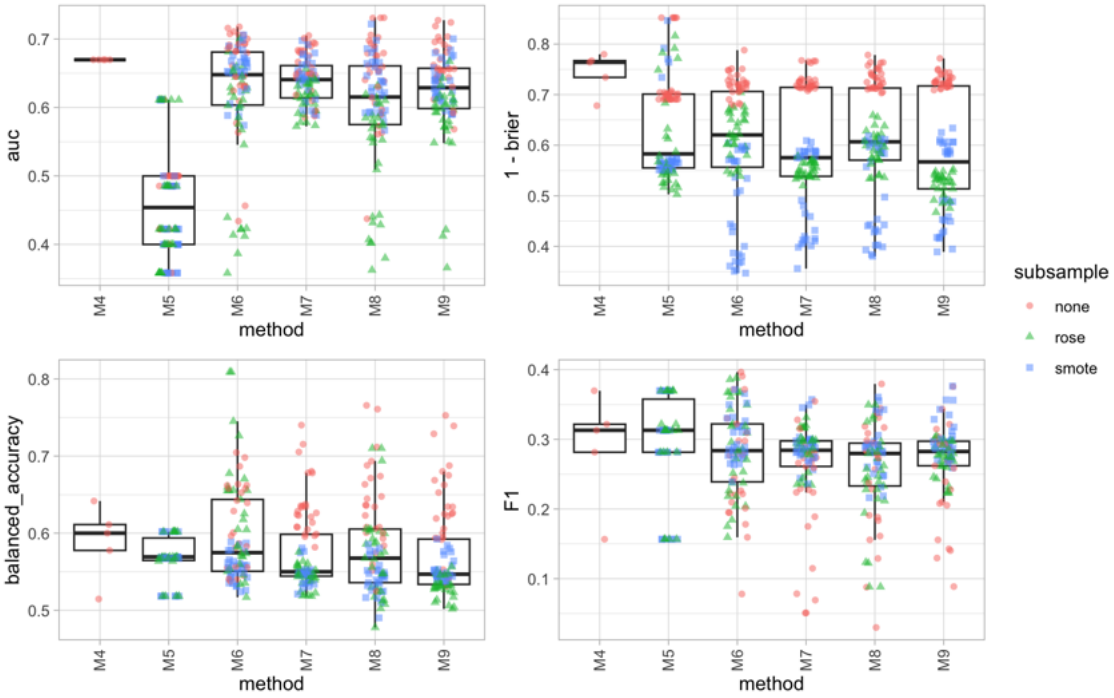


Figure 7.16: Aim 4 - Performance measures for object localization methods

The top five models, ranked by their performance scores after filtering out those that exceeded the 0.85 quantile for both the Brier score and AUC, are presented in the Table 7.4. When comparing balanced accuracy, where there's a baseline of 50% from always predicting negative, the model achieving the highest balanced accuracy is 71.2%. Regarding AUC, a range of 0.7 to 0.8 is generally deemed acceptable [66]. Additionally, we are not presenting

the best balanced accuracy or discussing the overall best balanced accuracy. Models that performed poorly in terms of AUC and Brier score have been excluded, potentially performing well in terms of balanced accuracy at the expense of a subpar performance in other aspects. In addition, all the five models are based on non-subsampled data.

It's noteworthy that all models in the top five are based on non-subsampled data, with XGBTree and XGBLinear models demonstrating the best performance. The α value of 0.6 outperforms the rest, suggesting the existence of an optimal blending weight for predicting situational awareness regarding object localization.

Table 7.4: Aim 4 - Top Five Performance Results of All Predictive Methods for Object Localization

M.α	ss.	model	TP	FP	FN	TN	acc.	b-acc	F1	prc.	rec.	AUC	Brier	prfm.
M8α.6	none	xgbTr	42	146	35	1048	0.858	0.712	0.317	0.223	0.545	0.685	0.266	0.514
M9α.6	none	xgbTr	51	137	58	1025	0.847	0.675	0.343	0.271	0.468	0.719	0.272	0.509
M6α.6	none	xgbTr	50	138	65	1018	0.840	0.658	0.330	0.266	0.435	0.711	0.271	0.494
M9αall	none	xgbLn	45	143	53	1030	0.846	0.669	0.315	0.239	0.459	0.727	0.228	0.492
M7αall	none	xgbLn	64	124	109	974	0.817	0.629	0.355	0.340	0.370	0.701	0.271	0.492

The highest-performing model for each method is recorded in Table 7.5. Method 5 notably exhibits inferior AUC, balanced accuracy, and F1 score compared to the others. Models with $\alpha = 0.6$ outshine other α values in half of the methods, where estimated gaze-point-based features and VA density are utilized. XGBTree also demonstrates superior performance across these three methods.

Table 7.5: Aim 4 - Top Performance Results of Each Predictive Method for Object Localization

M.α	ss.	model	TP	FP	FN	TN	acc.	b-acc	F1	prc.	rec.	AUC	Brier	prfm.
M4α0	none	rule	88	200	100	883	0.764	0.642	0.370	0.306	0.468	0.670	0.236	0.506
M5α.9	smote	rForest	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.208	0.338
M6α0	none	dt_gini	69	119	91	992	0.835	0.662	0.397	0.367	0.431	0.661	0.285	0.529
M7α.6	none	xgbTr	37	151	39	1044	0.851	0.680	0.280	0.197	0.487	0.680	0.269	0.480
M8α.6	none	xgbTr	42	146	35	1048	0.858	0.712	0.317	0.223	0.545	0.685	0.266	0.514
M9α.6	none	xgbTr	51	137	58	1025	0.847	0.675	0.343	0.271	0.468	0.719	0.272	0.509

(2) Object Recognition

The four primary performance metrics of all methods are illustrated in Figure 7.17. Notably, the Brier score has been adjusted to "1 - Brier score" to maintain consistency across the measures, ensuring that a value closer to 1 signifies better model performance.

Models utilizing SMOTE and ROSE subsampling methods demonstrate enhanced performance compared to non-sampled data in terms of AUC and Brier score. Conversely, models based on non-subsampled data excel in terms of balanced accuracy and F1 score. This underscores the effectiveness of our approach, which involves initially filtering out models based on AUC and Brier score, followed by comparing the model performance score, resulting in the models in the final pool performing better across all dimensions.

Method 4 excels in AUC, Brier score, and balanced accuracy but significantly lags behind the rest of the methods in terms of F1 score. Method 8 slightly outperforms in terms of AUC and balanced accuracy, while Method 6 performs marginally better regarding Brier score. All methods except Method 4 perform evenly regarding F1 score.

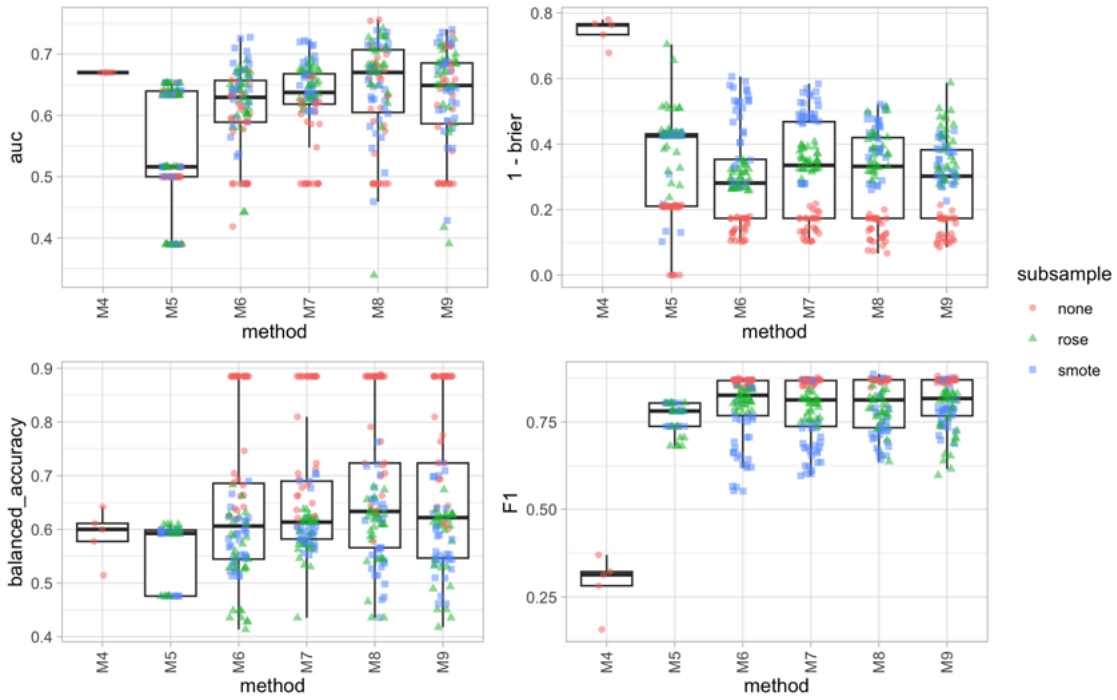


Figure 7.17: Aim 4 - Performance measures for object recognition methods

The top five models, ranked by their performance scores after filtering out those that exceeded the 0.85 quantile for both the Brier score and AUC, are presented in the Table 7.6. When comparing balanced accuracy, where there's a baseline of 50% from always predicting negative, the model achieving the highest balanced accuracy is 68.0%. Regarding AUC, a range of 0.7 to 0.8 is typically considered acceptable [66]. F1 scores above 0.7 or 0.8 are often considered good in many practical scenarios [94]. Additionally, we are not displaying the best balanced accuracy or F1 score. Models that performed poorly in terms of AUC and Brier score have been filtered out, which may perform well in terms of balanced accuracy or F1 score at the cost of a poor performance for the rest. It is noteworthy that all the models in the top five are based on subsampled data.

Table 7.6: Aim 4 - Top Five Performance Results of All Predictive Methods for Object Recognition

M.α	ss.	model	TP	FP	FN	TN	acc.	b-acc	F1	prc.	rec.	AUC	Brier	prfm.
M8αall	rose	dt_gini	85	59	4	40	0.665	0.680	0.730	0.590	0.955	0.728	0.502	0.705
M8α.9	smote	xgbLn	92	52	9	35	0.676	0.657	0.751	0.639	0.911	0.730	0.524	0.704
M8α.9	smote	rForest	87	57	8	36	0.654	0.651	0.728	0.604	0.916	0.711	0.506	0.690
M9α0	rose	xgbLn	84	60	10	34	0.628	0.628	0.706	0.583	0.894	0.725	0.522	0.667
M9α.3	rose	xgbTr	82	62	12	32	0.606	0.606	0.689	0.569	0.872	0.694	0.514	0.648

The highest-performing model for each approach is documented in Table 7.7. Initially, the leading models, as elaborated in section 6.7.4, underwent filtering based on the 50% quantile concerning both the Brier score and AUC for each method. Subsequently, they were ranked by their performance score, with only the top model being included in the table.

Method 4 notably demonstrates the best Brier score but with an inferior F1 score, while Method 5 exhibits poor balanced accuracy. The remaining methods perform uniformly. Three out of the four methods from Method 6 to 9 have $\alpha = 0.9$. No specific model stands out prominently. Notably, four methods achieve an F1 score exceeding 0.8, which is considered excellent.

Table 7.7: Aim 4 - Top Performance Results of Each Predictive Method for Object Recognition

M.α	ss.	model	TP	FP	FN	TN	acc.	b-acc	F1	prc.	rec.	AUC	Brier	prfm.
M4α0	none	rule	88	200	100	883	0.764	0.642	0.370	0.306	0.468	0.670	0.236	0.506
M5α.3	smote	log_reg	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.568	0.668
M6α.9	smote	log_reg	131	13	30	14	0.771	0.666	0.859	0.910	0.814	0.725	0.713	0.763
M7α.9	rose	xgbTree	117	27	24	20	0.729	0.628	0.821	0.812	0.830	0.655	0.671	0.724
M8α.3	rose	rForest	119	25	25	19	0.734	0.629	0.826	0.826	0.826	0.727	0.614	0.728
M9α.9	smote	dt_gini	112	32	20	24	0.723	0.639	0.812	0.778	0.848	0.725	0.627	0.725

7.8 Discussion

In terms of object localization, the blending weight seems to have a greater impact on model performance than the method itself. Models with $\alpha = 0.6$ ranked among the top three, alongside Methods 8, 9, and 6. Conversely, for object recognition, Method 8 performed better than others, particularly with $\alpha = 0.9$. This suggests that for different phases of SA,

object localization or recognition, relies on memory or trajectory on a different scale. Given that a larger α represents a greater dependency on the trajectory, the results indicate that the object recognition required more trajectory info for a more precise prediction compared to object localization.

Regarding the methods, Method 8 outperforms the others in predicting object recognition, underscoring the importance of estimated gaze features in this phase. While Method 8 also holds an advantage in object localization, it's not as pronounced. Notably, all models in the final pool integrate environmental, object, and driver features.

The proposed methods demonstrate comparable performance with Aim 3 which uses real gaze-related features for predicting driver SA. Regarding the balanced accuracy for object localization, Aim 3's best result is 69.5%, and for Aim 4, the number is 71.2%. Regarding the balanced accuracy for object recognition, Aim 3's best result is 62.8%, which is 68% for Aim 4. The method proposed in Aim 4 even outperforms Aim 3.

The proposed methodologies exhibit similar performance to those of Aim 3, which utilizes actual gaze-related features to predict driver SA. In terms of balanced accuracy for object localization, the best result achieved by Aim 3 is 69.5%, whereas for Aim 4, it reaches 71.2%. Concerning balanced accuracy for object recognition, Aim 3 achieves a maximum of 62.8%, whereas Aim 4 attains 68%. Notably, the approach proposed in Aim 4 even surpasses the performance of Aim 3.

In comparing F1 scores, for object localization, Aim 3 achieves its best at 0.372, while Aim 4 achieves 0.355. For object recognition, Aim 3 achieves its best at 0.757, whereas Aim 4 achieves 0.751. Despite a slight variation between the two, the difference is minimal.

In terms of the subsampling technique, results from both Aim 3 and Aim 4 follow a similar pattern. Models that utilize non-subsampled data demonstrate better performance in object localization, while those based on subsampled examples excel in object recognition. This variance could be attributed to the smaller dataset size for object recognition. Subsampling effectively enlarges the dataset, thereby improving model performance.

Concerning the feature sets, all models in the final pool for Aim 3 and Aim 4 integrate environmental, object, and driver features. This underscores their importance in achieving accurate situational awareness prediction, irrespective of whether it pertains to object

localization or object recognition.

Chapter 8
GENERAL CONCLUSION

This chapter summarizes the overall findings and contributions of the dissertation. Limitations and future research will also be discussed.

8.1 Overall Summary

The overall goal of this dissertation is to measure and predict drivers' SA to selectively assist drivers to avoid having warnings that exceed their SA capacity. The objective is divided into four aims: 1) Develop a method for effectively capturing driver SA, 2) Identify potential predictors of driver SA, 3) Predict drivers' SA using eye-tracking data, and 4) Predict driver SA using real-world driving videos, where a lower-cost and practical approach is explored.

8.1.1 Summary of Aim 1

Aim 1 introduces a novel approach to gathering driver SA data, capturing both direct SA measures (SAGAT responses) and indirect SA measures (eye tracking data) within a single controlled study. Prior research utilized query-based SA assessments, which treated SA as a binary outcome through participant responses to specific scene elements. However, this method had limitations, particularly in scenarios with multiple potential hazards. This study improves upon existing methods by pinpointing the exact locations in the scene where participants observed road users. Data collected from the experiment include drivers' demographic information obtained via pre-experiment surveys, SA responses recorded as object coordinates and types for each scenario using a web-based touch recorder, and drivers' eye gaze data collected via an eye tracker.

8.1.2 Summary of Aim 2

The study follows the SA capture approach from Aim 1, aiming to assess SA using gathered data and understand the influence of environmental, object, and driver features on SA before training predictive models. This aids in feature selection for subsequent model training in Aims 3 and 4. Features other than gaze features are focused on in this chapter for three reasons: exploring more predictors, eye gaze features being more suitable for models using eye tracking data, and feature extraction procedures still being in testing phase. The proposed SA collection method indicates potential for measuring higher-level SA, comprehension, and prediction. Environmental factors like number of objects, object size, visual complexity, and roadway type significantly impact driver SA, suggesting their inclusion in predictive models.

8.1.3 Summary of Aim 3

Aim 3 introduces predictive methods for driver SA based on eye gaze metrics, environmental properties, and object properties. The methods follow a hierarchical structure, focusing on object localization and object recognition phases. Binary classification models are employed for each phase, with the same pool of predictors utilized. Overall, the proposed methods consistently outperform baseline models in both phases, emphasizing the importance of incorporating gaze-point-based and visual sensory-dependent features into SA predictive models. Models using non-subsampled data excel in object localization, while subsampled data enhances object recognition, likely due to dataset size differences. Baseline 3 performs similarly to the proposed method for object recognition, indicating the significance of environmental, object, and driver features for accurate object recognition prediction.

8.1.4 Summary of Aim 4

Moving to Aim 4, the aim shifts towards building SA predictive models using real-world driving videos, without relying on eye-tracking data. The performance of these models is compared with those developed in Aim 3, which utilized actual gaze-related features. Interestingly, the results indicate that the models developed in Aim 4 perform comparably

to or even slightly better than those in Aim 3 in terms of balanced accuracy of the best models. This suggests that it may be possible to predict SA accurately using only video data, without the need for eye-tracking equipment. Furthermore, the analysis reveals that the choice of data subsampling technique (subsampling vs. non-subsampling) significantly impacts model performance, with non-subsampled data generally performing better for object localization and subsampled data performing better for object recognition. The blending weight, especially at $\alpha = 0.6$, significantly influences object localization model performance. However, for object recognition, Method 8 performs better, especially with $\alpha = 0.9$. This indicates varying reliance on memory or trajectory for different SA phases. A higher α suggests greater dependency on trajectory, with object recognition relying more on trajectory information for precise predictions compared to object localization. Additionally, it's noted that all models in the final pool incorporate environmental, object, and driver features, underscoring the importance of these factors in accurately predicting SA.

8.2 Contribution and Publications

The content of Aim 1 and 2 has been published in the Human Factors and Ergonomics Society (HFES) annual conference [104], AutomativeUI (AutoUI) conference [80], and the International Journal of Human-Computer Interaction [103].

Predicting Driver Situation Awareness regarding Object Localization and Recognition Using Eye-Tracking Data. This paper will propose a driver SA predictive model using eye tracking data, environmental features, object features, and driver's demographic features, which include the work from Aim 3. This paper will be submitted to *Transportation Research Part F: Traffic Psychology and Behaviour*.

Predicting Driver Situation Awareness Using Real-World Driving Videos. In this paper, a method will be proposed for employing visual attention predictive models to estimate eye fixation and forecast driver SA based on this estimated eye fixation information. This paper will be submitted to *IEEE Transactions on Intelligent Transportation Systems*.

8.3 Limitations and Future Research

One limitation of the study is the reliance on only one model, DeepGazeIIE, for estimating the probability distribution of drivers' gaze points. Future research should explore different models and conduct comparisons, as this is essential for improving implementation. While DeepGazeIIE demonstrates satisfactory performance upon manual inspection of the output results, it is not specifically designed for driving conditions. Moreover, models specifically trained for driving conditions are currently outdated. Therefore, it would be valuable to utilize newly emerging models tailored for driving conditions and repeat Aim 4 of this study to compare the results.

Another limitation is that the individual feature importance for the models was not thoroughly assessed. This study compared the outcomes of methods with different feature sets to identify which sets significantly influence model performance. Implementing a more rigorous process for selecting high-importance features could potentially enhance the model results.

One future research Sami Park from the HFSM lab as well and I are working on is to learn driver's SA especially when they are interacting with pedestrians. The paper has been submitted to *IEEE Transactions on Intelligent Vehicles*. Sami mainly focuses on the topic and will elaborate in her dissertation. Here is the abstract: Understanding and evaluating drivers' situation awareness (SA) is important to understand alert prioritization. This study investigates the relationship between driving performance measures (speed, acceleration and brake usage, steering wheel and lane deviation), pedestrian interaction (direction, motion and location), and driver SA. To achieve this, a controlled study was conducted with 56 participants using a Balanced Incomplete Block Design, where each participant drove 18 out of 48 possible intersections in a driving simulator environment. The Situational Awareness Global Assessment Technique (SAGAT) method was used to assess drivers' SA. Mixed effects logit models were developed to examine the different SA levels (perception, comprehension, projection). The driving performance measures were aggregated across three time windows (1, 3, and 5 seconds). The findings show significant contributions from both driving performance measures and pedestrian interactions in predicting driver SA. More

specifically, a one-second time window was useful for predicting pedestrian direction and a three-second time window was best for predicting pedestrian location and intention to cross. The results indicate the importance of considering different time windows for predicting various levels of driver SA responses. These findings offer insights into factors that should be considered in driver SA predictive models.

In terms of implementation, these real-time prediction models consistently require significant computational power. With advancements in cloud computing, vehicles may connect to external computational resources instead of relying on onboard systems in the future. Additionally, wearable devices can be used to supplement eye-tracking and video data to improve model performance.

BIBLIOGRAPHY

[1] Cheryl a. Bolstad, Haydee. Cuevas, Jingjing. Wang-Costello, Mica R. Endsley, and Linda S. Angell. Measurement of situation awareness for automobile technologies of the future. *Performance metrics for assessing driver distraction: the quest for improved road safety*, 2010.

[2] Ravi Agrawal, Timothy J Wright, Siby Samuel, Shlomo Zilberstein, and Donald L Fisher. Effects of a change in environment on the minimum time to situation awareness in transfer of control scenarios. *Transportation research record*, 2663(1):126–133, 2017.

[3] 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.

[4] 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.

[5] David Beattie, Lynne Baillie, Martin Halvey, and Rod McCall. What's around the corner? enhancing driver awareness in autonomous vehicles via in-vehicle spatial auditory displays. In *Proceedings of the 8th nordic conference on human-computer interaction: fun, fast, foundational*, pages 189–198, 2014.

[6] Steven S. Beauchemin and John L. Barron. The computation of optical flow. *ACM computing surveys (CSUR)*, 27(3):433–466, 1995.

[7] 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.

[8] Avinoam Borowsky, David Shinar, and Yisrael Parmet. The relation between driving experience and recognition of road signs relative to their locations. *Human factors*, 50(2):173–182, 2008.

[9] James F Brennan and Keith A Houde. *History and systems of psychology*. Cambridge University Press, 2022.

[10] 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.

[11] Peter J Burt and Edward H Adelson. A multiresolution spline with application to image mosaics. *ACM Transactions on Graphics (TOG)*, 2(4):217–236, 1983.

[12] 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.

[13] Peter R Chapman and Geoffrey Underwood. Visual search of driving situations: Danger and experience. *Perception*, 27(8):951–964, 1998.

[14] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[15] 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.

[16] Venkata Rami Reddy Chirra, Srinivasulu Reddy Uyyala, and Venkata Krishna Kishore Kolli. Deep cnn: A machine learning approach for driver drowsiness detection based on eye state. *Rev. d'Intelligence Artif.*, 33(6):461–466, 2019.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

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

[22] Tao Deng, Fei Yan, and Hongmei Yan. Driving video fixation prediction model via spatio-temporal networks and attention gates. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021.

[23] Tao Deng, Hongmei Yan, Long Qin, Thuyen Ngo, and BS Manjunath. How do drivers allocate their potential attention? driving fixation prediction via convolutional neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 21(5):2146–2154, 2019.

[24] 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.

[25] 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.

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

[27] M. R. Endsley. Situation awareness measurement in individuals & teams. In *Human Factors & Ergonomics Society Workshop*, 2020.

[28] 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.

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

[30] Mica R Endsley. A taxonomy of situation awareness errors. *Human factors in aviation operations*, 3(2):287–292, 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. Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*, 9(1):4–32, 2015.

[34] Mica R Endsley. Situation awareness. *Handbook of human factors and ergonomics*, pages 434–455, 2021.

[35] 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.

[36] Donald L Fisher and David L Strayer. Modeling situation awareness and crash risk. *Annals of Advances in Automotive Medicine*, 58:33, 2014.

[37] 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. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 64, pages 1836–1840, CA, USA, 2020. SAGE Publications Sage CA: Los Angeles, CA.

[38] 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.

[39] Chao Gou, Yuchen Zhou, and Dan Li. Driver attention prediction based on convolution and transformers. *The Journal of Supercomputing*, 78(6):8268–8284, 2022.

[40] 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.

[41] 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.

[42] 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.

[43] Markus Hofbauer, Christopher B Kuhn, Lukas Püttner, Goran Petrovic, and Eckehard Steinbach. Measuring driver situation awareness using region-of-interest prediction and eye tracking. In *2020 IEEE International Symposium on Multimedia (ISM)*, pages 91–95. IEEE, 2020.

[44] Becky L Hooey, Brian F Gore, Christopher D Wickens, Shelly Scott-Nash, Connie Socash, Ellen Salud, and David C Foyle. Modeling pilot situation awareness. In *Human modelling in assisted transportation: Models, tools and risk methods*, pages 207–213. Springer, 2011.

[45] M.S. Horswill and F.P. McKenna. Chapter 9 drivers' hazard perception ability: Situation awareness on the road. pages 115–175. 2004.

[46] Qing Hou, Chengbo Ai, Michael Knodler, and Francis Tainter. A framework for quantifying the impacts of situational visual clutter on driving performance via video analysis and eye tracking. *Available at SSRN 4100346*, 2019.

[47] Zhongxu Hu, Chen Lv, Peng Hang, Chao Huang, and Yang Xing. Data-driven estimation of driver attention using calibration-free eye gaze and scene features. *IEEE Transactions on Industrial Electronics*, 69(2):1800–1808, 2021.

[48] Tao Huang and Rui Fu. Driver distraction detection based on the true driver's focus of attention. *IEEE Transactions on Intelligent Transportation Systems*, 23(10):19374–19386, 2022.

[49] Katarzyna A Hussey, Sarah E Hadyniak, and Robert J Johnston Jr. Patterning and development of photoreceptors in the human retina. *Frontiers in Cell and Developmental Biology*, page 795, 2022.

[50] A. Hamish Jamson, Natasha Merat, Oliver M.J. Carsten, and Frank C.H. Lai. Behavioural changes in drivers experiencing highly-automated vehicle control in varying traffic conditions. *Transportation Research Part C: Emerging Technologies*, 2013.

[51] Debra G Jones and Mica R Endsley. Sources of situation awareness errors in aviation. *Aviation, space, and environmental medicine*, 1996.

[52] 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.

[53] David B Kaber, Yulan Liang, Yu Zhang, Meghan L Rogers, and Shruti Gangakhedkar. Driver performance effects of simultaneous visual and cognitive distraction and adaptation behavior. *Transportation research part F: traffic psychology and behaviour*, 15(5):491–501, 2012.

[54] Byeongkeun Kang and Yeejin Lee. High-resolution neural network for driver visual attention prediction. *Sensors*, 20(7):2030, 2020.

[55] 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.

[56] 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. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 2099–2103, CA, USA, 2019. SAGE Publications Sage CA: Los Angeles, CA.

[57] 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.

[58] Matthias Kümmerer, Matthias Bethge, and Thomas SA Wallis. Deepgaze iii: Modeling free-viewing human scanpaths with deep learning. *Journal of Vision*, 22(5):7–7, 2022.

[59] Fahad Lateef, Mohamed Kas, and Yassine Ruichek. Saliency heat-map as visual attention for autonomous driving using generative adversarial network (gan). *IEEE Transactions on Intelligent Transportation Systems*, 23(6):5360–5373, 2021.

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

[61] Akis Linardos, Matthias Kümmerer, Ori Press, and Matthias Bethge. Deepgaze iie: Calibrated prediction in and out-of-domain for state-of-the-art saliency modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12919–12928, 2021.

[62] 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.

[63] 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.

[64] 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.

[65] Alexander Makrigiorgos, Ali Shafti, Alex Harston, Julien Gerard, and A Aldo Faisal. Human visual attention prediction boosts learning & performance of autonomous driving agents. *arXiv preprint arXiv:1909.05003*, 2019.

[66] Jayawant N Mandrekar. Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*, 5(9):1315–1316, 2010.

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

[68] Giovanna Menardi and Nicola Torelli. Training and assessing classification rules with imbalanced data. *Data mining and knowledge discovery*, 28:92–122, 2014.

[69] 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.

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

[71] David Moher, Alessandro Liberati, Jennifer Tetzlaff, Douglas G Altman, et al. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Int J Surg*, 8(5):336–341, 2010.

[72] 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.

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

[74] 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.

[75] Minghao Ning, Chao Lu, and Jianwei Gong. An efficient model for driving focus of attention prediction using deep learning. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 1192–1197. IEEE, 2019.

[76] Gary W Oehlert. *A first course in design and analysis of experiments*. W. H. Freeman and Company, 2010.

[77] Andrea Palazzi, Davide Abati, Francesco Solera, Rita Cucchiara, et al. Predicting the driver's focus of attention: the dr (eye) ve project. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1720–1733, 2018.

[78] 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.

[79] Daehee Park, Wan Chul Yoon, and Uichin Lee. Cognitive states matter: Design guidelines for driving situation awareness in smart vehicles. *Sensors (Switzerland)*, 2020.

[80] 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.

[81] 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.

[82] Hong Pei-heng and Yuehua Wang. Real-time driver's focus of attention extraction and prediction using deep learning. *International Journal of Advanced Computer Science and Applications*, 12(6), 2021.

[83] Long Qin, Yi Shi, Yahui He, Junrui Zhang, Xianshi Zhang, Yongjie Li, Tao Deng, and Hongmei Yan. Id-yolo: Real-time salient object detection based on the driver's fixation region. *IEEE Transactions on Intelligent Transportation Systems*, 23(9):15898–15908, 2022.

[84] Yuning Qiu, Carlos Busso, Teruhisa Misu, and Kumar Akash. Incorporating gaze behavior using joint embedding with scene context for driver takeover detection. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4633–4637. IEEE, 2022.

[85] 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.

[86] 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.

[87] Umair Rehman, Shi Cao, and Carolyn MacGregor. Using an integrated cognitive architecture to model the effect of environmental complexity on drivers' situation awareness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 63, pages 812–816, Los Angeles, CA, 2019. SAGE Publications Sage CA: Los Angeles, CA, SAGE Publications.

[88] 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.

[89] William B Rouse and Nancy M Morris. On looking into the black box: Prospects and limits in the search for mental models. *Psychological bulletin*, 100(3):349, 1986.

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

[91] 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.

[92] Mohsen Shirpour, Steven S Beauchemin, and Michael A Bauer. Driver's eye fixation prediction by deep neural network. In *VISIGRAPP (4: VISAPP)*, pages 67–75, 2021.

[93] R Jay Shively, Michael Brickner, and Jacob Silbiger. A computational model of situational awareness instantiated in midas(man-machine integration design and analysis). In *International Symposium on Aviation Psychology, 9 th, Columbus, OH*, pages 1454–1459, 1997.

[94] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437, 2009.

[95] 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.

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

[97] Richard M Taylor. Situational awareness rating technique (sart): The development of a tool for aircrew systems design. In *Situational awareness*, pages 111–128. Routledge, 2017.

[98] 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.

[99] Michael A Vidulich and Pamela S Tsang. Chapter 2.8 mental workload and situation awareness. *Handbook of Human Factors and Ergonomics: Fourth Edition*, pages 243–273, 2012.

[100] MinJuan Wang, Sus Lundgren Lyckvi, Chenhui Chen, Palle Dahlstedt, and Fang Chen. Using advisory 3d sound cues to improve drivers' performance and situation awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 2814–2825, 2017.

[101] 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.

[102] Christopher D Wickens, William S Helton, Justin G Hollands, and Simon Banbury. *Engineering psychology and human performance*. Routledge, 2012.

[103] Yilun Xing, Sami Park, Kumar Akash, Teruhisa Misu, 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.

[104] 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. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 66(1), pages 335–339. SAGE Publications Sage CA: Los Angeles, CA, 2022.

[105] Yanqun Yang, Meifeng Chen, Changxu Wu, Said M Easa, and Xinyi Zheng. Structural equation modeling of drivers' situation awareness considering road and driver factors. *Frontiers in psychology*, 11:1601, 2020.

[106] Kristie L Young, Amanda N Stephens, David B Logan, and Michael G Lenné. Investigating the impact of static roadside advertising on drivers' situation awareness. *Applied ergonomics*, 60:136–145, 2017.

[107] 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.

[108] 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 Transportation Systems*, 23(3):2284–2295, 2021.

[109] 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.

Appendix A

MODEL PERFORMANCE MEASURES

A.1 Aim 3

A.1.1 Object Localization

Table A.1: Model Performance Measures for Object Localization of Aim 3

	method	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	Brier	prfm.
1	Baseline1	none	rule-based	53	140	135	943	0.784	0.576	0.278	0.275	0.282	0.670	0.216	0.427
2	Baseline2	none	logistic_reg	0	188	0	1083	0.852			0.000		0.424	0.299	
3	Baseline2	none	decision_tree_gini	0	188	0	1083	0.852			0.000		0.500	0.309	
4	Baseline2	none	decision_tree_entropy	0	188	0	1083	0.852			0.000		0.500	0.309	
5	Baseline2	none	random_forest	0	188	0	1083	0.852			0.000		0.500	0.148	
6	Baseline2	none	xgbTree	0	188	0	1083	0.852			0.000		0.424	0.295	
7	Baseline2	none	xgbLinear	0	188	0	1083	0.852			0.000		0.424	0.299	
8	Baseline2	smote	logistic_reg	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.434	0.426
9	Baseline2	smote	decision_tree_gini	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.434	0.426
10	Baseline2	smote	decision_tree_entropy	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.434	0.426
11	Baseline2	smote	random_forest	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.216	0.426
12	Baseline2	smote	xgbTree	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.434	0.426
13	Baseline2	smote	xgbLinear	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.434	0.426
14	Baseline2	rose	logistic_reg	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.483	0.426
15	Baseline2	rose	decision_tree_gini	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.437	0.426

Table A.1: Model Performance Measures for Object Localization of Aim 3

	method	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	Brier	prfm.
16	Baseline2	rose	decision_tree_entropy	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.437	0.426
17	Baseline2	rose	random_forest	0	188	0	1083	0.852			0.000		0.424	0.165	
18	Baseline2	rose	xgbTree	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.424	0.404	0.426
19	Baseline2	rose	xgbLinear	0	188	0	1083	0.852			0.000		0.424	0.311	
20	Baseline3	none	logistic_reg	0	188	1	1082	0.851	0.426		0.000	0.000	0.622	0.280	
21	Baseline3	none	decision_tree_gini	33	155	85	998	0.811	0.573	0.216	0.176	0.280	0.658	0.279	0.394
22	Baseline3	none	decision_tree_entropy	40	148	83	1000	0.818	0.598	0.257	0.213	0.325	0.658	0.281	0.428
23	Baseline3	none	random_forest	12	176	10	1073	0.854	0.702	0.114	0.064	0.545	0.690	0.209	0.408
24	Baseline3	none	xgbTree	50	138	64	1019	0.841	0.660	0.331	0.266	0.439	0.689	0.269	0.495
25	Baseline3	none	xgbLinear	51	137	113	970	0.803	0.594	0.290	0.271	0.311	0.697	0.268	0.442
26	Baseline3	smote	logistic_reg	59	129	188	895	0.751	0.556	0.271	0.314	0.239	0.613	0.405	0.414
27	Baseline3	smote	decision_tree_gini	84	104	268	815	0.707	0.563	0.311	0.447	0.239	0.646	0.396	0.437
28	Baseline3	smote	decision_tree_entropy	127	61	353	730	0.674	0.594	0.380	0.676	0.265	0.670	0.417	0.487
29	Baseline3	smote	random_forest	165	23	835	248	0.325	0.540	0.278	0.878	0.165	0.679	0.589	0.409
30	Baseline3	smote	xgbTree	153	35	767	316	0.369	0.533	0.276	0.814	0.166	0.695	0.571	0.405
31	Baseline3	smote	xgbLinear	175	13	913	170	0.271	0.545	0.274	0.931	0.161	0.696	0.675	0.410
32	Baseline3	rose	logistic_reg	72	116	276	807	0.692	0.541	0.269	0.383	0.207	0.614	0.452	0.405
33	Baseline3	rose	decision_tree_gini	53	135	118	965	0.801	0.594	0.295	0.282	0.310	0.628	0.385	0.444
34	Baseline3	rose	decision_tree_entropy	53	135	128	955	0.793	0.584	0.287	0.282	0.293	0.374	0.390	0.436
35	Baseline3	rose	random_forest	43	145	84	999	0.820	0.606	0.273	0.229	0.339	0.642	0.337	0.439
36	Baseline3	rose	xgbTree	32	156	29	1054	0.854	0.698	0.257	0.170	0.525	0.642	0.316	0.477
37	Baseline3	rose	xgbLinear	45	143	88	995	0.818	0.606	0.280	0.239	0.338	0.597	0.299	0.443
38	M1	none	logistic_reg	0	188	0	1083	0.852			0.000		0.607	0.287	
39	M1	none	decision_tree_gini	30	158	59	1024	0.829	0.602	0.217	0.160	0.337	0.655	0.275	0.409
40	M1	none	decision_tree_entropy	33	155	46	1037	0.842	0.644	0.247	0.176	0.418	0.672	0.272	0.446
41	M1	none	random_forest	39	149	67	1016	0.830	0.620	0.265	0.207	0.368	0.734	0.282	0.443
42	M1	none	xgbTree	56	132	69	1014	0.842	0.666	0.358	0.298	0.448	0.733	0.243	0.512

Table A.1: Model Performance Measures for Object Localization of Aim 3

	method	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	Brier	prfm.
43	M1	none	xgbLinear	49	139	87	996	0.822	0.619	0.302	0.261	0.360	0.727	0.241	0.461
44	M1	smote	logistic_reg	66	122	249	834	0.708	0.541	0.262	0.351	0.210	0.603	0.409	0.402
45	M1	smote	decision_tree_gini	118	70	279	804	0.725	0.609	0.403	0.628	0.297	0.696	0.369	0.506
46	M1	smote	decision_tree_entropy	118	70	279	804	0.725	0.609	0.403	0.628	0.297	0.701	0.369	0.506
47	M1	smote	random_forest	152	36	617	466	0.486	0.563	0.318	0.809	0.198	0.704	0.512	0.440
48	M1	smote	xgbTree	151	37	640	443	0.467	0.557	0.308	0.803	0.191	0.704	0.515	0.433
49	M1	smote	xgbLinear	152	36	659	424	0.453	0.555	0.304	0.809	0.187	0.703	0.531	0.429
50	M1	rose	logistic_reg	89	99	384	699	0.620	0.532	0.269	0.473	0.188	0.594	0.463	0.401
51	M1	rose	decision_tree_gini	22	166	8	1075	0.863	0.800	0.202	0.117	0.733	0.392	0.417	0.501
52	M1	rose	decision_tree_entropy	22	166	8	1075	0.863	0.800	0.202	0.117	0.733	0.394	0.421	0.501
53	M1	rose	random_forest	38	150	102	981	0.802	0.569	0.232	0.202	0.271	0.615	0.406	0.401
54	M1	rose	xgbTree	29	159	59	1024	0.828	0.598	0.210	0.154	0.330	0.647	0.404	0.404
55	M1	rose	xgbLinear	43	145	118	965	0.793	0.568	0.246	0.229	0.267	0.581	0.376	0.407
56	M2	none	logistic_reg	0	188	0	1083	0.852			0.000		0.601	0.287	
57	M2	none	decision_tree_gini	39	149	84	999	0.817	0.594	0.251	0.207	0.317	0.655	0.281	0.422
58	M2	none	decision_tree_entropy	37	151	75	1008	0.822	0.600	0.247	0.197	0.330	0.660	0.279	0.423
59	M2	none	random_forest	40	148	58	1025	0.838	0.641	0.280	0.213	0.408	0.735	0.287	0.460
60	M2	none	xgbTree	40	148	39	1044	0.853	0.691	0.300	0.213	0.506	0.766	0.256	0.495
61	M2	none	xgbLinear	65	123	96	987	0.828	0.646	0.372	0.346	0.404	0.755	0.240	0.509
62	M2	smote	logistic_reg	66	122	251	832	0.707	0.540	0.261	0.351	0.208	0.602	0.409	0.401
63	M2	smote	decision_tree_gini	106	82	308	775	0.693	0.580	0.352	0.564	0.256	0.659	0.389	0.466
64	M2	smote	decision_tree_entropy	109	79	316	767	0.689	0.582	0.356	0.580	0.256	0.670	0.382	0.469
65	M2	smote	random_forest	147	41	513	570	0.564	0.578	0.347	0.782	0.223	0.720	0.463	0.462
66	M2	smote	xgbTree	145	43	537	546	0.544	0.570	0.333	0.771	0.213	0.717	0.473	0.452
67	M2	smote	xgbLinear	160	28	621	462	0.489	0.574	0.330	0.851	0.205	0.713	0.510	0.452
68	M2	rose	logistic_reg	74	114	342	741	0.641	0.522	0.245	0.394	0.178	0.587	0.457	0.384
69	M2	rose	decision_tree_gini	15	173	8	1075	0.858	0.757	0.142	0.080	0.652	0.422	0.419	0.449

Table A.1: Model Performance Measures for Object Localization of Aim 3

	method	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	Brier	prfm.
70	M2	rose	decision_tree_entropy	9	179	7	1076	0.854	0.710	0.088	0.048	0.562	0.390	0.417	0.399
71	M2	rose	random_forest	76	112	260	823	0.707	0.553	0.290	0.404	0.226	0.615	0.431	0.422
72	M2	rose	xgbTree	74	114	276	807	0.693	0.544	0.275	0.394	0.211	0.601	0.428	0.409
73	M2	rose	xgbLinear	84	104	464	619	0.553	0.505	0.228	0.447	0.153	0.527	0.446	0.366
74	M3	none	logistic_reg	0	188	0	1083	0.852			0.000		0.602	0.287	
75	M3	none	decision_tree_gini	37	151	71	1012	0.825	0.606	0.250	0.197	0.343	0.653	0.281	0.428
76	M3	none	decision_tree_entropy	42	146	88	995	0.816	0.598	0.264	0.223	0.323	0.656	0.279	0.431
77	M3	none	random_forest	40	148	56	1027	0.839	0.645	0.282	0.213	0.417	0.748	0.284	0.464
78	M3	none	xgbTree	43	145	42	1041	0.853	0.692	0.315	0.229	0.506	0.751	0.253	0.503
79	M3	none	xgbLinear	42	146	80	1003	0.822	0.609	0.271	0.223	0.344	0.719	0.238	0.440
80	M3	smote	logistic_reg	71	117	267	816	0.698	0.542	0.270	0.378	0.210	0.600	0.407	0.406
81	M3	smote	decision_tree_gini	118	70	307	776	0.703	0.597	0.385	0.628	0.278	0.680	0.385	0.491
82	M3	smote	decision_tree_entropy	120	68	316	767	0.698	0.597	0.385	0.638	0.275	0.693	0.380	0.491
83	M3	smote	random_forest	132	56	422	661	0.624	0.580	0.356	0.702	0.238	0.705	0.460	0.468
84	M3	smote	xgbTree	136	52	514	569	0.555	0.563	0.325	0.723	0.209	0.698	0.458	0.444
85	M3	smote	xgbLinear	152	36	507	576	0.573	0.586	0.359	0.809	0.231	0.722	0.450	0.472
86	M3	rose	logistic_reg	83	105	363	720	0.632	0.529	0.262	0.441	0.186	0.589	0.461	0.396
87	M3	rose	decision_tree_gini	19	169	8	1075	0.861	0.784	0.177	0.101	0.704	0.407	0.422	0.480
88	M3	rose	decision_tree_entropy	19	169	8	1075	0.861	0.784	0.177	0.101	0.704	0.407	0.422	0.480
89	M3	rose	random_forest	96	92	345	738	0.656	0.553	0.305	0.511	0.218	0.614	0.445	0.429
90	M3	rose	xgbTree	62	126	175	908	0.763	0.570	0.292	0.330	0.262	0.624	0.403	0.431
91	M3	rose	xgbLinear	59	129	228	855	0.719	0.537	0.248	0.314	0.206	0.585	0.340	0.393

A.1.2 Object Recognition

Table A.2: Model Performance Measures for Object Recognition of Aim 3

	method	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	Brier	prfm.
1	Baseline1	none	rule-based	40	13	104	31	0.378	0.491	0.406	0.755	0.278	0.670	0.622	0.449
2	Baseline2	none	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.790	
3	Baseline2	none	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
4	Baseline2	none	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
5	Baseline2	none	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.500	1.000	
6	Baseline2	none	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.793	
7	Baseline2	none	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.790	
8	Baseline2	smote	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.573	
9	Baseline2	smote	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.571	
10	Baseline2	smote	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.571	
11	Baseline2	smote	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.996	
12	Baseline2	smote	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.569	
13	Baseline2	smote	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.573	
14	Baseline2	rose	logistic_reg	104	40	31	13	0.622	0.508	0.746	0.722	0.770	0.491	0.491	0.627
15	Baseline2	rose	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.626	
16	Baseline2	rose	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.491	0.580	
17	Baseline2	rose	random_forest	40	104	13	31	0.378	0.492	0.406	0.278	0.755	0.509	0.524	0.449
18	Baseline2	rose	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.509	0.606	
19	Baseline2	rose	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.509	0.647	
20	Baseline3	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.664	0.895	0.878
21	Baseline3	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
22	Baseline3	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
23	Baseline3	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.599	0.860	0.878
24	Baseline3	none	xgbTree	134	10	35	9	0.761	0.633	0.856	0.931	0.793	0.550	0.846	0.745
25	Baseline3	none	xgbLinear	130	14	30	14	0.766	0.656	0.855	0.903	0.812	0.581	0.824	0.756

Table A.2: Model Performance Measures for Object Recognition of Aim 3

	method	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	Brier	prfm.
26	Baseline3	smote	logistic_reg	131	13	27	17	0.787	0.698	0.868	0.910	0.829	0.713	0.722	0.783
27	Baseline3	smote	decision_tree_gini	83	61	18	26	0.580	0.560	0.678	0.576	0.822	0.659	0.466	0.619
28	Baseline3	smote	decision_tree_entropy	106	38	26	18	0.660	0.562	0.768	0.736	0.803	0.636	0.584	0.665
29	Baseline3	smote	random_forest	55	89	6	38	0.495	0.600	0.537	0.382	0.902	0.682	0.434	0.569
30	Baseline3	smote	xgbTree	72	72	10	34	0.564	0.599	0.637	0.500	0.878	0.629	0.450	0.618
31	Baseline3	smote	xgbLinear	69	75	10	34	0.548	0.593	0.619	0.479	0.873	0.662	0.430	0.606
32	Baseline3	rose	logistic_reg	117	27	25	19	0.723	0.618	0.818	0.812	0.824	0.671	0.674	0.718
33	Baseline3	rose	decision_tree_gini	112	32	23	21	0.707	0.613	0.803	0.778	0.830	0.666	0.662	0.708
34	Baseline3	rose	decision_tree_entropy	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.687	0.639
35	Baseline3	rose	random_forest	117	27	24	20	0.729	0.628	0.821	0.812	0.830	0.671	0.652	0.724
36	Baseline3	rose	xgbTree	115	29	28	16	0.697	0.580	0.801	0.799	0.804	0.669	0.671	0.691
37	Baseline3	rose	xgbLinear	120	24	27	17	0.729	0.615	0.825	0.833	0.816	0.697	0.736	0.720
38	M1	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.685	0.894	0.878
39	M1	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
40	M1	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
41	M1	none	random_forest	137	7	39	5	0.755	0.598	0.856	0.951	0.778	0.664	0.771	0.727
42	M1	none	xgbTree	143	1	42	2	0.771	0.720	0.869	0.993	0.773	0.692	0.899	0.795
43	M1	none	xgbLinear	133	11	35	9	0.755	0.621	0.853	0.924	0.792	0.643	0.845	0.737
44	M1	smote	logistic_reg	134	10	32	12	0.777	0.676	0.865	0.931	0.807	0.728	0.715	0.770
45	M1	smote	decision_tree_gini	73	71	9	35	0.574	0.610	0.646	0.507	0.890	0.682	0.455	0.628
46	M1	smote	decision_tree_entropy	96	48	18	26	0.649	0.597	0.744	0.667	0.842	0.674	0.571	0.670
47	M1	smote	random_forest	77	67	7	37	0.606	0.636	0.675	0.535	0.917	0.682	0.501	0.656
48	M1	smote	xgbTree	77	67	15	29	0.564	0.570	0.653	0.535	0.837	0.655	0.507	0.611
49	M1	smote	xgbLinear	77	67	13	31	0.574	0.586	0.658	0.535	0.856	0.630	0.480	0.622
50	M1	rose	logistic_reg	128	16	31	13	0.750	0.627	0.845	0.889	0.805	0.681	0.696	0.736
51	M1	rose	decision_tree_gini	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.683	0.639
52	M1	rose	decision_tree_entropy	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.683	0.639

Table A.2: Model Performance Measures for Object Recognition of Aim 3

	method	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	Brier	prfm.
53	M1	rose	random_forest	114	30	23	21	0.718	0.622	0.811	0.792	0.832	0.700	0.617	0.717
54	M1	rose	xgbTree	108	36	21	23	0.697	0.614	0.791	0.750	0.837	0.711	0.620	0.702
55	M1	rose	xgbLinear	118	26	18	26	0.766	0.684	0.843	0.819	0.868	0.741	0.691	0.763
56	M2	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.678	0.895	0.878
57	M2	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
58	M2	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
59	M2	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.670	0.821	0.878
60	M2	none	xgbTree	143	1	41	3	0.777	0.764	0.872	0.993	0.777	0.630	0.880	0.818
61	M2	none	xgbLinear	134	10	36	8	0.755	0.616	0.854	0.931	0.788	0.623	0.863	0.735
62	M2	smote	logistic_reg	136	8	31	13	0.793	0.717	0.875	0.944	0.814	0.736	0.731	0.796
63	M2	smote	decision_tree_gini	106	38	27	17	0.654	0.553	0.765	0.736	0.797	0.645	0.580	0.659
64	M2	smote	decision_tree_entropy	115	29	31	13	0.681	0.549	0.793	0.799	0.788	0.650	0.667	0.671
65	M2	smote	random_forest	74	70	12	32	0.564	0.587	0.643	0.514	0.860	0.638	0.504	0.615
66	M2	smote	xgbTree	85	59	16	28	0.601	0.582	0.694	0.590	0.842	0.656	0.547	0.638
67	M2	smote	xgbLinear	83	61	17	27	0.585	0.568	0.680	0.576	0.830	0.634	0.534	0.624
68	M2	rose	logistic_reg	129	15	31	13	0.755	0.635	0.849	0.896	0.806	0.668	0.700	0.742
69	M2	rose	decision_tree_gini	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.686	0.639
70	M2	rose	decision_tree_entropy	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.658	0.639
71	M2	rose	random_forest	117	27	26	18	0.718	0.609	0.815	0.812	0.818	0.669	0.629	0.712
72	M2	rose	xgbTree	115	29	23	21	0.723	0.627	0.816	0.799	0.833	0.673	0.645	0.721
73	M2	rose	xgbLinear	120	24	27	17	0.729	0.615	0.825	0.833	0.816	0.694	0.719	0.720
74	M3	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.673	0.894	0.878
75	M3	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
76	M3	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
77	M3	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.682	0.808	0.878
78	M3	none	xgbTree	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.691	0.854	0.878
79	M3	none	xgbLinear	132	12	37	7	0.739	0.575	0.843	0.917	0.781	0.674	0.860	0.709

Table A.2: Model Performance Measures for Object Recognition of Aim 3

	method	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	Brier	prfm.
80	M3	smote	logistic_reg	126	18	30	14	0.745	0.623	0.840	0.875	0.808	0.704	0.719	0.731
81	M3	smote	decision_tree_gini	74	70	13	31	0.559	0.579	0.641	0.514	0.851	0.649	0.445	0.610
82	M3	smote	decision_tree_entropy	108	36	24	20	0.681	0.588	0.783	0.750	0.818	0.584	0.596	0.685
83	M3	smote	random_forest	81	63	11	33	0.606	0.612	0.686	0.562	0.880	0.683	0.530	0.649
84	M3	smote	xgbTree	98	46	17	27	0.665	0.611	0.757	0.681	0.852	0.711	0.593	0.684
85	M3	smote	xgbLinear	85	59	14	30	0.612	0.598	0.700	0.590	0.859	0.684	0.538	0.649
86	M3	rose	logistic_reg	122	22	29	15	0.729	0.607	0.827	0.847	0.808	0.661	0.682	0.717
87	M3	rose	decision_tree_gini	121	23	28	16	0.729	0.611	0.826	0.840	0.812	0.608	0.684	0.719
88	M3	rose	decision_tree_entropy	121	23	28	16	0.729	0.611	0.826	0.840	0.812	0.542	0.705	0.719
89	M3	rose	random_forest	113	31	22	22	0.718	0.626	0.810	0.785	0.837	0.670	0.595	0.718
90	M3	rose	xgbTree	108	36	22	22	0.691	0.605	0.788	0.750	0.831	0.655	0.636	0.697
91	M3	rose	xgbLinear	114	30	26	18	0.702	0.595	0.803	0.792	0.814	0.657	0.689	0.699

A.2 Aim 4

A.2.1 Object Localization

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
1	M4	a0	none	rule-based	88	200	100	883	0.764	0.642	0.370	0.306	0.468	0.670	0.236	0.506
2	M4	a3	none	rule-based	77	227	111	856	0.734	0.600	0.313	0.253	0.410	0.670	0.266	0.456
3	M4	a6	none	rule-based	58	166	130	917	0.767	0.578	0.282	0.259	0.309	0.670	0.233	0.430
4	M4	a9	none	rule-based	26	118	162	965	0.780	0.515	0.157	0.181	0.138	0.670	0.220	0.336
5	M4	all	none	rule-based	97	318	91	765	0.678	0.611	0.322	0.234	0.516	0.670	0.322	0.466
6	M5	a0	none	logistic_reg	0	188	0	1083	0.852			0.000		0.358	0.295	
7	M5	a0	none	decision_tree_gini	0	188	0	1083	0.852			0.000		0.500	0.309	
8	M5	a0	none	decision_tree_entropy	0	188	0	1083	0.852			0.000		0.500	0.309	
9	M5	a0	none	random_forest	0	188	0	1083	0.852			0.000		0.500	0.148	
10	M5	a0	none	xgbTree	0	188	0	1083	0.852			0.000		0.358	0.292	
11	M5	a0	none	xgbLinear	0	188	0	1083	0.852			0.000		0.358	0.295	
12	M5	a0	smote	logistic_reg	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.427	0.486
13	M5	a0	smote	decision_tree_gini	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.427	0.486
14	M5	a0	smote	decision_tree_entropy	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.427	0.486
15	M5	a0	smote	random_forest	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.236	0.486
16	M5	a0	smote	xgbTree	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.427	0.486
17	M5	a0	smote	xgbLinear	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.427	0.486
18	M5	a0	rose	logistic_reg	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.476	0.486
19	M5	a0	rose	decision_tree_gini	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.447	0.486
20	M5	a0	rose	decision_tree_entropy	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.447	0.486
21	M5	a0	rose	random_forest	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.228	0.486
22	M5	a0	rose	xgbTree	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.421	0.486

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
23	M5	a0	rose	xgbLinear	88	100	200	883	0.764	0.602	0.370	0.468	0.306	0.358	0.368	0.486
24	M5	a3	none	logistic_reg	0	188	0	1083	0.852			0.000		0.400	0.300	
25	M5	a3	none	decision_tree_gini	0	188	0	1083	0.852			0.000		0.500	0.309	
26	M5	a3	none	decision_tree_entropy	0	188	0	1083	0.852			0.000		0.500	0.309	
27	M5	a3	none	random_forest	0	188	0	1083	0.852			0.000		0.500	0.148	
28	M5	a3	none	xgbTree	0	188	0	1083	0.852			0.000		0.400	0.297	
29	M5	a3	none	xgbLinear	0	188	0	1083	0.852			0.000		0.400	0.300	
30	M5	a3	smote	logistic_reg	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.435	0.441
31	M5	a3	smote	decision_tree_gini	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.435	0.441
32	M5	a3	smote	decision_tree_entropy	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.435	0.441
33	M5	a3	smote	random_forest	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.266	0.441
34	M5	a3	smote	xgbTree	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.435	0.441
35	M5	a3	smote	xgbLinear	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.435	0.441
36	M5	a3	rose	logistic_reg	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.484	0.441
37	M5	a3	rose	decision_tree_gini	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.459	0.441
38	M5	a3	rose	decision_tree_entropy	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.459	0.441
39	M5	a3	rose	random_forest	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.251	0.441
40	M5	a3	rose	xgbTree	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.436	0.441
41	M5	a3	rose	xgbLinear	77	111	227	856	0.734	0.569	0.313	0.410	0.253	0.400	0.413	0.441
42	M5	a6	none	logistic_reg	0	188	0	1083	0.852			0.000		0.422	0.302	
43	M5	a6	none	decision_tree_gini	0	188	0	1083	0.852			0.000		0.500	0.309	
44	M5	a6	none	decision_tree_entropy	0	188	0	1083	0.852			0.000		0.500	0.309	
45	M5	a6	none	random_forest	0	188	0	1083	0.852			0.000		0.500	0.148	
46	M5	a6	none	xgbTree	0	188	0	1083	0.852			0.000		0.422	0.298	
47	M5	a6	none	xgbLinear	0	188	0	1083	0.852			0.000		0.422	0.302	
48	M5	a6	smote	logistic_reg	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.439	0.424
49	M5	a6	smote	decision_tree_gini	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.439	0.424

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
50	M5	a6	smote	decision_tree_entropy	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.439	0.424
51	M5	a6	smote	random_forest	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.230	0.424
52	M5	a6	smote	xgbTree	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.438	0.424
53	M5	a6	smote	xgbLinear	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.439	0.424
54	M5	a6	rose	logistic_reg	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.488	0.424
55	M5	a6	rose	decision_tree_gini	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.475	0.424
56	M5	a6	rose	decision_tree_entropy	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.475	0.424
57	M5	a6	rose	random_forest	0	188	0	1083	0.852			0.000		0.422	0.184	
58	M5	a6	rose	xgbTree	58	130	166	917	0.767	0.567	0.282	0.309	0.259	0.422	0.439	0.424
59	M5	a6	rose	xgbLinear	0	188	0	1083	0.852			0.000		0.422	0.383	
60	M5	a9	none	logistic_reg	0	188	0	1083	0.852			0.000		0.485	0.309	
61	M5	a9	none	decision_tree_gini	0	188	0	1083	0.852			0.000		0.500	0.309	
62	M5	a9	none	decision_tree_entropy	0	188	0	1083	0.852			0.000		0.500	0.309	
63	M5	a9	none	random_forest	0	188	0	1083	0.852			0.000		0.500	0.148	
64	M5	a9	none	xgbTree	0	188	0	1083	0.852			0.000		0.485	0.305	
65	M5	a9	none	xgbLinear	0	188	0	1083	0.852			0.000		0.485	0.309	
66	M5	a9	smote	logistic_reg	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.449	0.338
67	M5	a9	smote	decision_tree_gini	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.449	0.338
68	M5	a9	smote	decision_tree_entropy	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.449	0.338
69	M5	a9	smote	random_forest	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.208	0.338
70	M5	a9	smote	xgbTree	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.448	0.338
71	M5	a9	smote	xgbLinear	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.449	0.338
72	M5	a9	rose	logistic_reg	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.497	0.338
73	M5	a9	rose	decision_tree_gini	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.452	0.338
74	M5	a9	rose	decision_tree_entropy	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.455	0.338
75	M5	a9	rose	random_forest	26	162	118	965	0.780	0.518	0.157	0.138	0.181	0.485	0.217	0.338
76	M5	a9	rose	xgbTree	0	188	0	1083	0.852			0.000		0.485	0.433	

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
77	M5	a9	rose	xgbLinear	0	188	0	1083	0.852			0.000		0.485	0.317	
78	M5	all	none	logistic_reg	0	188	0	1083	0.852			0.000		0.611	0.298	
79	M5	all	none	decision_tree_gini	0	188	0	1083	0.852			0.000		0.500	0.309	
80	M5	all	none	decision_tree_entropy	0	188	0	1083	0.852			0.000		0.500	0.309	
81	M5	all	none	random_forest	0	188	0	1083	0.852			0.000		0.500	0.148	
82	M5	all	none	xgbTree	0	188	0	1083	0.852			0.000		0.611	0.296	
83	M5	all	none	xgbLinear	0	188	0	1083	0.852			0.000		0.611	0.298	
84	M5	all	smote	logistic_reg	0	188	0	1083	0.852			0.000		0.611	0.432	
85	M5	all	smote	decision_tree_gini	0	188	0	1083	0.852			0.000		0.500	0.450	
86	M5	all	smote	decision_tree_entropy	0	188	0	1083	0.852			0.000		0.500	0.450	
87	M5	all	smote	random_forest	0	188	0	1083	0.852			0.000		0.611	0.153	
88	M5	all	smote	xgbTree	0	188	0	1083	0.852			0.000		0.611	0.432	
89	M5	all	smote	xgbLinear	0	188	0	1083	0.852			0.000		0.611	0.432	
90	M5	all	rose	logistic_reg	97	91	318	765	0.678	0.564	0.322	0.516	0.234	0.611	0.483	0.443
91	M5	all	rose	decision_tree_gini	97	91	318	765	0.678	0.564	0.322	0.516	0.234	0.611	0.469	0.443
92	M5	all	rose	decision_tree_entropy	0	188	0	1083	0.852			0.000		0.611	0.453	
93	M5	all	rose	random_forest	0	188	0	1083	0.852			0.000		0.611	0.209	
94	M5	all	rose	xgbTree	0	188	0	1083	0.852			0.000		0.611	0.456	
95	M5	all	rose	xgbLinear	0	188	0	1083	0.852			0.000		0.611	0.387	
96	M6	a0	none	logistic_reg	0	188	0	1083	0.852			0.000		0.680	0.275	
97	M6	a0	none	decision_tree_gini	69	119	91	992	0.835	0.662	0.397	0.367	0.431	0.661	0.285	0.529
98	M6	a0	none	decision_tree_entropy	59	129	149	934	0.781	0.581	0.298	0.314	0.284	0.564	0.322	0.440
99	M6	a0	none	random_forest	30	158	27	1056	0.854	0.698	0.245	0.160	0.526	0.691	0.249	0.471
100	M6	a0	none	xgbTree	36	152	36	1047	0.852	0.687	0.277	0.191	0.500	0.700	0.265	0.482
101	M6	a0	none	xgbLinear	49	139	115	968	0.800	0.587	0.278	0.261	0.299	0.706	0.261	0.433
102	M6	a0	smote	logistic_reg	74	114	184	899	0.766	0.587	0.332	0.394	0.287	0.651	0.395	0.459
103	M6	a0	smote	decision_tree_gini	114	74	414	669	0.616	0.558	0.318	0.606	0.216	0.606	0.463	0.438

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
104	M6	a0	smote	decision_tree_entropy	114	74	414	669	0.616	0.558	0.318	0.606	0.216	0.588	0.466	0.438
105	M6	a0	smote	random_forest	169	19	867	216	0.303	0.541	0.276	0.899	0.163	0.671	0.615	0.409
106	M6	a0	smote	xgbTree	153	35	758	325	0.376	0.535	0.278	0.814	0.168	0.669	0.561	0.407
107	M6	a0	smote	xgbLinear	179	9	893	190	0.290	0.561	0.284	0.952	0.167	0.697	0.653	0.423
108	M6	a0	rose	logistic_reg	89	99	242	841	0.732	0.582	0.343	0.473	0.269	0.681	0.447	0.462
109	M6	a0	rose	decision_tree_gini	61	127	83	1000	0.835	0.655	0.367	0.324	0.424	0.422	0.399	0.511
110	M6	a0	rose	decision_tree_entropy	61	127	83	1000	0.835	0.655	0.367	0.324	0.424	0.422	0.399	0.511
111	M6	a0	rose	random_forest	66	122	88	995	0.835	0.660	0.386	0.351	0.429	0.700	0.337	0.523
112	M6	a0	rose	xgbTree	59	129	80	1003	0.836	0.655	0.361	0.314	0.424	0.637	0.338	0.508
113	M6	a0	rose	xgbLinear	64	124	78	1005	0.841	0.670	0.388	0.340	0.451	0.660	0.296	0.529
114	M6	a3	none	logistic_reg	0	188	0	1083	0.852			0.000		0.632	0.280	
115	M6	a3	none	decision_tree_gini	38	150	90	993	0.811	0.583	0.241	0.202	0.297	0.434	0.295	0.412
116	M6	a3	none	decision_tree_entropy	43	145	149	934	0.769	0.545	0.226	0.229	0.224	0.456	0.320	0.386
117	M6	a3	none	random_forest	8	180	9	1074	0.851	0.664	0.078	0.043	0.471	0.686	0.212	0.371
118	M6	a3	none	xgbTree	67	121	88	995	0.836	0.662	0.391	0.356	0.432	0.706	0.280	0.526
119	M6	a3	none	xgbLinear	53	135	140	943	0.784	0.575	0.278	0.282	0.275	0.693	0.288	0.426
120	M6	a3	smote	logistic_reg	57	131	188	895	0.749	0.552	0.263	0.303	0.233	0.623	0.402	0.408
121	M6	a3	smote	decision_tree_gini	114	74	350	733	0.666	0.577	0.350	0.606	0.246	0.687	0.409	0.463
122	M6	a3	smote	decision_tree_entropy	121	67	471	612	0.577	0.553	0.310	0.644	0.204	0.587	0.490	0.432
123	M6	a3	smote	random_forest	168	20	911	172	0.268	0.526	0.265	0.894	0.156	0.676	0.631	0.395
124	M6	a3	smote	xgbTree	171	17	846	237	0.321	0.551	0.284	0.910	0.168	0.667	0.633	0.417
125	M6	a3	smote	xgbLinear	170	18	870	213	0.301	0.543	0.277	0.904	0.163	0.674	0.646	0.410
126	M6	a3	rose	logistic_reg	76	112	267	816	0.702	0.550	0.286	0.404	0.222	0.629	0.452	0.418
127	M6	a3	rose	decision_tree_gini	61	127	83	1000	0.835	0.655	0.367	0.324	0.424	0.358	0.387	0.511
128	M6	a3	rose	decision_tree_entropy	61	127	83	1000	0.835	0.655	0.367	0.324	0.424	0.387	0.385	0.511
129	M6	a3	rose	random_forest	46	142	77	1006	0.828	0.625	0.296	0.245	0.374	0.660	0.350	0.460
130	M6	a3	rose	xgbTree	43	145	47	1036	0.849	0.678	0.309	0.229	0.478	0.645	0.351	0.493

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
131	M6	a3	rose	xgbLinear	33	155	57	1026	0.833	0.618	0.237	0.176	0.367	0.624	0.326	0.428
132	M6	a6	none	logistic_reg	0	188	0	1083	0.852			0.000		0.609	0.283	
133	M6	a6	none	decision_tree_gini	22	166	66	1017	0.817	0.555	0.159	0.117	0.250	0.645	0.291	0.357
134	M6	a6	none	decision_tree_entropy	39	149	145	938	0.769	0.537	0.210	0.207	0.212	0.632	0.317	0.374
135	M6	a6	none	random_forest	24	164	34	1049	0.844	0.639	0.195	0.128	0.414	0.693	0.251	0.417
136	M6	a6	none	xgbTree	50	138	65	1018	0.840	0.658	0.330	0.266	0.435	0.711	0.271	0.494
137	M6	a6	none	xgbLinear	56	132	118	965	0.803	0.601	0.309	0.298	0.322	0.692	0.277	0.455
138	M6	a6	smote	logistic_reg	69	119	220	863	0.733	0.559	0.289	0.367	0.239	0.599	0.409	0.424
139	M6	a6	smote	decision_tree_gini	113	75	383	700	0.640	0.566	0.330	0.601	0.228	0.650	0.433	0.448
140	M6	a6	smote	decision_tree_entropy	128	60	384	699	0.651	0.585	0.366	0.681	0.250	0.651	0.431	0.476
141	M6	a6	smote	random_forest	156	32	799	284	0.346	0.531	0.273	0.830	0.163	0.653	0.571	0.402
142	M6	a6	smote	xgbTree	152	36	712	371	0.411	0.544	0.289	0.809	0.176	0.690	0.556	0.416
143	M6	a6	smote	xgbLinear	172	16	889	194	0.288	0.543	0.275	0.915	0.162	0.684	0.649	0.409
144	M6	a6	rose	logistic_reg	80	108	297	786	0.681	0.546	0.283	0.426	0.212	0.599	0.457	0.414
145	M6	a6	rose	decision_tree_gini	28	160	20	1063	0.858	0.726	0.237	0.149	0.583	0.414	0.378	0.482
146	M6	a6	rose	decision_tree_entropy	58	130	86	997	0.830	0.644	0.349	0.309	0.403	0.412	0.404	0.497
147	M6	a6	rose	random_forest	33	155	102	981	0.798	0.554	0.204	0.176	0.244	0.657	0.356	0.379
148	M6	a6	rose	xgbTree	41	147	26	1057	0.864	0.745	0.322	0.218	0.612	0.629	0.333	0.533
149	M6	a6	rose	xgbLinear	24	164	89	994	0.801	0.535	0.159	0.128	0.212	0.546	0.324	0.347
150	M6	a9	none	logistic_reg	0	188	0	1083	0.852			0.000		0.577	0.287	
151	M6	a9	none	decision_tree_gini	26	162	78	1005	0.811	0.556	0.178	0.138	0.250	0.645	0.293	0.367
152	M6	a9	none	decision_tree_entropy	43	145	157	926	0.762	0.540	0.222	0.229	0.215	0.631	0.319	0.381
153	M6	a9	none	random_forest	28	160	36	1047	0.846	0.652	0.222	0.149	0.438	0.682	0.249	0.437
154	M6	a9	none	xgbTree	66	122	101	982	0.825	0.642	0.372	0.351	0.395	0.718	0.284	0.507
155	M6	a9	none	xgbLinear	71	117	123	960	0.811	0.629	0.372	0.378	0.366	0.716	0.280	0.500
156	M6	a9	smote	logistic_reg	57	131	232	851	0.714	0.532	0.239	0.303	0.197	0.574	0.409	0.385
157	M6	a9	smote	decision_tree_gini	104	84	376	707	0.638	0.555	0.311	0.553	0.217	0.646	0.402	0.433

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
158	M6	a9	smote	decision_tree_entropy	127	61	369	714	0.662	0.589	0.371	0.676	0.256	0.665	0.427	0.480
159	M6	a9	smote	random_forest	165	23	893	190	0.279	0.524	0.265	0.878	0.156	0.659	0.619	0.394
160	M6	a9	smote	xgbTree	148	40	594	489	0.501	0.562	0.318	0.787	0.199	0.706	0.494	0.440
161	M6	a9	smote	xgbLinear	169	19	797	286	0.358	0.556	0.293	0.899	0.175	0.699	0.613	0.425
162	M6	a9	rose	logistic_reg	62	126	298	785	0.666	0.517	0.226	0.330	0.172	0.568	0.461	0.372
163	M6	a9	rose	decision_tree_gini	38	150	74	1009	0.824	0.605	0.253	0.202	0.339	0.439	0.398	0.429
164	M6	a9	rose	decision_tree_entropy	43	145	85	998	0.819	0.605	0.272	0.229	0.336	0.423	0.397	0.438
165	M6	a9	rose	random_forest	29	159	67	1016	0.822	0.583	0.204	0.154	0.302	0.622	0.349	0.394
166	M6	a9	rose	xgbTree	24	164	8	1075	0.865	0.809	0.218	0.128	0.750	0.585	0.365	0.513
167	M6	a9	rose	xgbLinear	27	161	78	1005	0.812	0.560	0.184	0.144	0.257	0.576	0.334	0.372
168	M6	all	none	logistic_reg	0	188	0	1083	0.852			0.000		0.624	0.280	
169	M6	all	none	decision_tree_gini	30	158	90	993	0.805	0.556	0.195	0.160	0.250	0.645	0.295	0.376
170	M6	all	none	decision_tree_entropy	35	153	125	958	0.781	0.541	0.201	0.186	0.219	0.629	0.309	0.371
171	M6	all	none	random_forest	26	162	22	1061	0.855	0.705	0.220	0.138	0.542	0.682	0.253	0.462
172	M6	all	none	xgbTree	48	140	62	1021	0.841	0.658	0.322	0.255	0.436	0.707	0.277	0.490
173	M6	all	none	xgbLinear	53	135	115	968	0.803	0.597	0.298	0.282	0.315	0.688	0.272	0.447
174	M6	all	smote	logistic_reg	67	121	177	906	0.766	0.578	0.310	0.356	0.275	0.627	0.405	0.444
175	M6	all	smote	decision_tree_gini	126	62	402	681	0.635	0.578	0.352	0.670	0.239	0.668	0.432	0.465
176	M6	all	smote	decision_tree_entropy	111	77	381	702	0.640	0.563	0.326	0.590	0.226	0.659	0.429	0.445
177	M6	all	smote	random_forest	169	19	890	193	0.285	0.535	0.271	0.899	0.160	0.652	0.630	0.403
178	M6	all	smote	xgbTree	167	21	814	269	0.343	0.549	0.286	0.888	0.170	0.666	0.600	0.417
179	M6	all	smote	xgbLinear	171	17	856	227	0.313	0.548	0.281	0.910	0.167	0.653	0.639	0.415
180	M6	all	rose	logistic_reg	65	123	264	819	0.696	0.533	0.251	0.346	0.198	0.615	0.453	0.392
181	M6	all	rose	decision_tree_gini	77	111	218	865	0.741	0.574	0.319	0.410	0.261	0.603	0.420	0.446
182	M6	all	rose	decision_tree_entropy	77	111	218	865	0.741	0.574	0.319	0.410	0.261	0.603	0.420	0.446
183	M6	all	rose	random_forest	52	136	144	939	0.780	0.569	0.271	0.277	0.265	0.653	0.380	0.420
184	M6	all	rose	xgbTree	24	164	8	1075	0.865	0.809	0.218	0.128	0.750	0.646	0.379	0.513

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
185	M6	all	rose	xgbLinear	24	164	62	1021	0.822	0.570	0.175	0.128	0.279	0.611	0.329	0.373
186	M7	a0	none	logistic_reg	0	188	0	1083	0.852			0.000		0.585	0.288	
187	M7	a0	none	decision_tree_gini	31	157	58	1025	0.831	0.608	0.224	0.165	0.348	0.644	0.286	0.416
188	M7	a0	none	decision_tree_entropy	38	150	58	1025	0.836	0.634	0.268	0.202	0.396	0.666	0.275	0.451
189	M7	a0	none	random_forest	8	180	8	1075	0.852	0.678	0.078	0.043	0.500	0.693	0.234	0.378
190	M7	a0	none	xgbTree	24	164	42	1041	0.838	0.614	0.189	0.128	0.364	0.695	0.274	0.401
191	M7	a0	none	xgbLinear	62	126	128	955	0.800	0.605	0.328	0.330	0.326	0.681	0.281	0.466
192	M7	a0	smote	logistic_reg	62	126	258	825	0.698	0.531	0.244	0.330	0.194	0.589	0.414	0.387
193	M7	a0	smote	decision_tree_gini	107	81	411	672	0.613	0.549	0.303	0.569	0.207	0.647	0.411	0.426
194	M7	a0	smote	decision_tree_entropy	120	68	364	719	0.660	0.581	0.357	0.638	0.248	0.665	0.422	0.469
195	M7	a0	smote	random_forest	165	23	817	266	0.339	0.544	0.282	0.878	0.168	0.642	0.595	0.413
196	M7	a0	smote	xgbTree	146	42	659	424	0.448	0.546	0.294	0.777	0.181	0.677	0.520	0.420
197	M7	a0	smote	xgbLinear	159	29	724	359	0.408	0.553	0.297	0.846	0.180	0.680	0.580	0.425
198	M7	a0	rose	logistic_reg	68	120	323	760	0.651	0.519	0.235	0.362	0.174	0.572	0.462	0.377
199	M7	a0	rose	decision_tree_gini	130	58	566	517	0.509	0.543	0.294	0.691	0.187	0.615	0.480	0.419
200	M7	a0	rose	decision_tree_entropy	94	94	382	701	0.625	0.540	0.283	0.500	0.197	0.614	0.467	0.411
201	M7	a0	rose	random_forest	93	95	344	739	0.655	0.549	0.298	0.495	0.213	0.623	0.435	0.424
202	M7	a0	rose	xgbTree	102	86	439	644	0.587	0.535	0.280	0.543	0.189	0.605	0.461	0.408
203	M7	a0	rose	xgbLinear	85	103	321	762	0.666	0.545	0.286	0.452	0.209	0.634	0.432	0.416
204	M7	a3	none	logistic_reg	0	188	0	1083	0.852			0.000		0.588	0.287	
205	M7	a3	none	decision_tree_gini	39	149	59	1024	0.836	0.635	0.273	0.207	0.398	0.648	0.282	0.454
206	M7	a3	none	decision_tree_entropy	33	155	67	1016	0.825	0.599	0.229	0.176	0.330	0.650	0.282	0.414
207	M7	a3	none	random_forest	5	183	4	1079	0.853	0.705	0.051	0.027	0.556	0.683	0.232	0.378
208	M7	a3	none	xgbTree	30	158	50	1033	0.836	0.621	0.224	0.160	0.375	0.658	0.280	0.423
209	M7	a3	none	xgbLinear	51	137	110	973	0.806	0.597	0.292	0.271	0.317	0.664	0.274	0.444
210	M7	a3	smote	logistic_reg	65	123	254	829	0.703	0.537	0.256	0.346	0.204	0.591	0.413	0.397
211	M7	a3	smote	decision_tree_gini	104	84	407	676	0.614	0.546	0.298	0.553	0.204	0.644	0.413	0.422

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
212	M7	a3	smote	decision_tree_entropy	107	81	420	663	0.606	0.547	0.299	0.569	0.203	0.644	0.412	0.423
213	M7	a3	smote	random_forest	166	22	814	269	0.342	0.547	0.284	0.883	0.169	0.649	0.592	0.416
214	M7	a3	smote	xgbTree	147	41	661	422	0.448	0.547	0.295	0.782	0.182	0.698	0.535	0.421
215	M7	a3	smote	xgbLinear	160	28	763	320	0.378	0.546	0.288	0.851	0.173	0.654	0.602	0.417
216	M7	a3	rose	logistic_reg	68	120	322	761	0.652	0.519	0.235	0.362	0.174	0.574	0.462	0.377
217	M7	a3	rose	decision_tree_gini	75	113	249	834	0.715	0.556	0.293	0.399	0.231	0.621	0.456	0.425
218	M7	a3	rose	decision_tree_entropy	94	94	385	698	0.623	0.539	0.282	0.500	0.196	0.613	0.464	0.410
219	M7	a3	rose	random_forest	88	100	315	768	0.673	0.552	0.298	0.468	0.218	0.619	0.429	0.425
220	M7	a3	rose	xgbTree	72	116	258	825	0.706	0.547	0.278	0.383	0.218	0.630	0.436	0.413
221	M7	a3	rose	xgbLinear	81	107	373	710	0.622	0.524	0.252	0.431	0.178	0.589	0.449	0.388
222	M7	a6	none	logistic_reg	0	188	0	1083	0.852			0.000		0.595	0.287	
223	M7	a6	none	decision_tree_gini	39	149	58	1025	0.837	0.638	0.274	0.207	0.402	0.649	0.284	0.456
224	M7	a6	none	decision_tree_entropy	42	146	69	1014	0.831	0.626	0.281	0.223	0.378	0.648	0.283	0.454
225	M7	a6	none	random_forest	7	181	7	1076	0.852	0.678	0.069	0.037	0.500	0.689	0.234	0.374
226	M7	a6	none	xgbTree	37	151	39	1044	0.851	0.680	0.280	0.197	0.487	0.680	0.269	0.480
227	M7	a6	none	xgbLinear	52	136	89	994	0.823	0.624	0.316	0.277	0.369	0.676	0.272	0.470
228	M7	a6	smote	logistic_reg	71	117	257	826	0.706	0.546	0.275	0.378	0.216	0.596	0.411	0.411
229	M7	a6	smote	decision_tree_gini	110	78	441	642	0.592	0.546	0.298	0.585	0.200	0.641	0.422	0.422
230	M7	a6	smote	decision_tree_entropy	114	74	358	725	0.660	0.574	0.345	0.606	0.242	0.662	0.425	0.460
231	M7	a6	smote	random_forest	161	27	769	314	0.374	0.547	0.288	0.856	0.173	0.640	0.586	0.417
232	M7	a6	smote	xgbTree	145	43	631	452	0.470	0.550	0.301	0.771	0.187	0.673	0.509	0.425
233	M7	a6	smote	xgbLinear	153	35	751	332	0.382	0.537	0.280	0.814	0.169	0.653	0.589	0.409
234	M7	a6	rose	logistic_reg	68	120	324	759	0.651	0.518	0.234	0.362	0.173	0.578	0.461	0.376
235	M7	a6	rose	decision_tree_gini	97	91	327	756	0.671	0.561	0.317	0.516	0.229	0.640	0.451	0.439
236	M7	a6	rose	decision_tree_entropy	97	91	327	756	0.671	0.561	0.317	0.516	0.229	0.639	0.451	0.439
237	M7	a6	rose	random_forest	85	103	296	787	0.686	0.554	0.299	0.452	0.223	0.637	0.435	0.426
238	M7	a6	rose	xgbTree	86	102	245	838	0.727	0.576	0.331	0.457	0.260	0.656	0.422	0.454

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
239	M7	a6	rose	xgbLinear	94	94	366	717	0.638	0.544	0.290	0.500	0.204	0.606	0.431	0.417
240	M7	a9	none	logistic_reg	0	188	0	1083	0.852			0.000		0.610	0.285	
241	M7	a9	none	decision_tree_gini	39	149	74	1009	0.825	0.608	0.259	0.207	0.345	0.624	0.285	0.434
242	M7	a9	none	decision_tree_entropy	41	147	85	998	0.817	0.599	0.261	0.218	0.325	0.618	0.289	0.430
243	M7	a9	none	random_forest	5	183	3	1080	0.854	0.740	0.051	0.027	0.625	0.683	0.237	0.396
244	M7	a9	none	xgbTree	21	167	31	1052	0.844	0.633	0.175	0.112	0.404	0.705	0.269	0.404
245	M7	a9	none	xgbLinear	46	142	112	971	0.800	0.582	0.266	0.245	0.291	0.652	0.276	0.424
246	M7	a9	smote	logistic_reg	77	111	265	818	0.704	0.553	0.291	0.410	0.225	0.608	0.407	0.422
247	M7	a9	smote	decision_tree_gini	66	122	305	778	0.664	0.521	0.236	0.351	0.178	0.612	0.394	0.379
248	M7	a9	smote	decision_tree_entropy	79	109	325	758	0.659	0.535	0.267	0.420	0.196	0.625	0.391	0.401
249	M7	a9	smote	random_forest	165	23	789	294	0.361	0.550	0.289	0.878	0.173	0.633	0.584	0.420
250	M7	a9	smote	xgbTree	143	45	687	396	0.424	0.535	0.281	0.761	0.172	0.644	0.541	0.408
251	M7	a9	smote	xgbLinear	158	30	726	357	0.405	0.551	0.295	0.840	0.179	0.692	0.568	0.423
252	M7	a9	rose	logistic_reg	73	115	336	747	0.645	0.523	0.245	0.388	0.178	0.591	0.459	0.384
253	M7	a9	rose	decision_tree_gini	122	66	497	586	0.557	0.548	0.302	0.649	0.197	0.636	0.463	0.425
254	M7	a9	rose	decision_tree_entropy	107	81	456	627	0.577	0.538	0.285	0.569	0.190	0.595	0.455	0.411
255	M7	a9	rose	random_forest	110	78	394	689	0.629	0.558	0.318	0.585	0.218	0.644	0.446	0.438
256	M7	a9	rose	xgbTree	114	74	419	664	0.612	0.557	0.316	0.606	0.214	0.640	0.444	0.437
257	M7	a9	rose	xgbLinear	116	72	476	607	0.569	0.545	0.297	0.617	0.196	0.618	0.453	0.421
258	M7	all	none	logistic_reg	0	188	0	1083	0.852			0.000		0.595	0.287	
259	M7	all	none	decision_tree_gini	39	149	59	1024	0.836	0.635	0.273	0.207	0.398	0.648	0.282	0.454
260	M7	all	none	decision_tree_entropy	33	155	69	1014	0.824	0.595	0.228	0.176	0.324	0.649	0.283	0.412
261	M7	all	none	random_forest	12	176	9	1074	0.854	0.715	0.115	0.064	0.571	0.695	0.235	0.415
262	M7	all	none	xgbTree	52	136	90	993	0.822	0.623	0.315	0.277	0.366	0.684	0.293	0.469
263	M7	all	none	xgbLinear	64	124	109	974	0.817	0.629	0.355	0.340	0.370	0.701	0.271	0.492
264	M7	all	smote	logistic_reg	72	116	260	823	0.704	0.547	0.277	0.383	0.217	0.595	0.412	0.412
265	M7	all	smote	decision_tree_gini	110	78	441	642	0.592	0.546	0.298	0.585	0.200	0.641	0.422	0.422

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
266	M7	all	smote	decision_tree_entropy	116	72	360	723	0.660	0.577	0.349	0.617	0.244	0.662	0.424	0.463
267	M7	all	smote	random_forest	165	23	796	287	0.356	0.549	0.287	0.878	0.172	0.649	0.587	0.418
268	M7	all	smote	xgbTree	157	31	817	266	0.333	0.528	0.270	0.835	0.161	0.678	0.599	0.399
269	M7	all	smote	xgbLinear	163	25	843	240	0.317	0.534	0.273	0.867	0.162	0.639	0.644	0.403
270	M7	all	rose	logistic_reg	69	119	324	759	0.651	0.520	0.238	0.367	0.176	0.578	0.462	0.379
271	M7	all	rose	decision_tree_gini	94	94	306	777	0.685	0.564	0.320	0.500	0.235	0.634	0.436	0.442
272	M7	all	rose	decision_tree_entropy	107	81	425	658	0.602	0.546	0.297	0.569	0.201	0.620	0.463	0.421
273	M7	all	rose	random_forest	90	98	323	760	0.669	0.552	0.300	0.479	0.218	0.631	0.440	0.426
274	M7	all	rose	xgbTree	79	109	302	781	0.677	0.542	0.278	0.420	0.207	0.619	0.430	0.410
275	M7	all	rose	xgbLinear	90	98	349	734	0.648	0.544	0.287	0.479	0.205	0.605	0.435	0.415
276	M8	a0	none	logistic_reg	0	188	0	1083	0.852			0.000		0.606	0.282	
277	M8	a0	none	decision_tree_gini	58	130	118	965	0.805	0.605	0.319	0.309	0.330	0.651	0.283	0.462
278	M8	a0	none	decision_tree_entropy	54	134	84	999	0.828	0.637	0.331	0.287	0.391	0.676	0.272	0.484
279	M8	a0	none	random_forest	9	179	8	1075	0.853	0.693	0.088	0.048	0.529	0.640	0.257	0.391
280	M8	a0	none	xgbTree	45	143	117	966	0.795	0.574	0.257	0.239	0.278	0.639	0.287	0.416
281	M8	a0	none	xgbLinear	48	140	75	1008	0.831	0.634	0.309	0.255	0.390	0.706	0.249	0.471
282	M8	a0	smote	logistic_reg	63	125	259	824	0.698	0.532	0.247	0.335	0.196	0.604	0.392	0.390
283	M8	a0	smote	decision_tree_gini	75	113	341	742	0.643	0.524	0.248	0.399	0.180	0.593	0.415	0.386
284	M8	a0	smote	decision_tree_entropy	85	103	299	784	0.684	0.553	0.297	0.452	0.221	0.639	0.396	0.425
285	M8	a0	smote	random_forest	166	22	921	162	0.258	0.517	0.260	0.883	0.153	0.624	0.607	0.388
286	M8	a0	smote	xgbTree	145	43	701	382	0.415	0.535	0.280	0.771	0.171	0.626	0.557	0.408
287	M8	a0	smote	xgbLinear	164	24	773	310	0.373	0.552	0.292	0.872	0.175	0.646	0.597	0.422
288	M8	a0	rose	logistic_reg	63	125	302	781	0.664	0.517	0.228	0.335	0.173	0.589	0.460	0.373
289	M8	a0	rose	decision_tree_gini	9	179	7	1076	0.854	0.710	0.088	0.048	0.562	0.440	0.378	0.399
290	M8	a0	rose	decision_tree_entropy	56	132	152	931	0.777	0.573	0.283	0.298	0.269	0.407	0.384	0.428
291	M8	a0	rose	random_forest	59	129	247	836	0.704	0.530	0.239	0.314	0.193	0.569	0.426	0.384
292	M8	a0	rose	xgbTree	37	151	80	1003	0.818	0.593	0.243	0.197	0.316	0.591	0.360	0.418

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
293	M8	a0	rose	xgbLinear	55	133	229	854	0.715	0.529	0.233	0.293	0.194	0.581	0.392	0.381
294	M8	a3	none	logistic_reg	0	188	0	1083	0.852			0.000		0.563	0.288	
295	M8	a3	none	decision_tree_gini	52	136	116	967	0.802	0.593	0.292	0.277	0.310	0.653	0.288	0.443
296	M8	a3	none	decision_tree_entropy	47	141	97	986	0.813	0.601	0.283	0.250	0.326	0.437	0.285	0.442
297	M8	a3	none	random_forest	3	185	9	1074	0.847	0.552	0.030	0.016	0.250	0.679	0.242	0.291
298	M8	a3	none	xgbTree	22	166	63	1020	0.820	0.559	0.161	0.117	0.259	0.660	0.260	0.360
299	M8	a3	none	xgbLinear	34	154	64	1019	0.828	0.608	0.238	0.181	0.347	0.679	0.250	0.423
300	M8	a3	smote	logistic_reg	58	130	255	828	0.697	0.525	0.232	0.309	0.185	0.579	0.409	0.378
301	M8	a3	smote	decision_tree_gini	86	102	330	753	0.660	0.544	0.285	0.457	0.207	0.589	0.432	0.414
302	M8	a3	smote	decision_tree_entropy	91	97	341	742	0.655	0.548	0.294	0.484	0.211	0.596	0.403	0.421
303	M8	a3	smote	random_forest	168	20	896	187	0.279	0.531	0.268	0.894	0.158	0.616	0.619	0.400
304	M8	a3	smote	xgbTree	153	35	740	343	0.390	0.539	0.283	0.814	0.171	0.638	0.567	0.411
305	M8	a3	smote	xgbLinear	169	19	825	258	0.336	0.551	0.286	0.899	0.170	0.629	0.616	0.418
306	M8	a3	rose	logistic_reg	62	126	328	755	0.643	0.508	0.215	0.330	0.159	0.556	0.465	0.361
307	M8	a3	rose	decision_tree_gini	74	114	182	901	0.767	0.588	0.333	0.394	0.289	0.363	0.409	0.461
308	M8	a3	rose	decision_tree_entropy	80	108	224	859	0.739	0.576	0.325	0.426	0.263	0.632	0.423	0.450
309	M8	a3	rose	random_forest	48	140	98	985	0.813	0.602	0.287	0.255	0.329	0.614	0.396	0.445
310	M8	a3	rose	xgbTree	39	149	78	1005	0.821	0.602	0.256	0.207	0.333	0.548	0.372	0.429
311	M8	a3	rose	xgbLinear	41	147	114	969	0.795	0.566	0.239	0.218	0.265	0.607	0.344	0.403
312	M8	a6	none	logistic_reg	0	188	0	1083	0.852			0.000		0.573	0.288	
313	M8	a6	none	decision_tree_gini	56	132	88	995	0.827	0.636	0.337	0.298	0.389	0.654	0.289	0.487
314	M8	a6	none	decision_tree_entropy	52	136	126	957	0.794	0.584	0.284	0.277	0.292	0.615	0.295	0.434
315	M8	a6	none	random_forest	20	168	10	1073	0.860	0.766	0.183	0.106	0.667	0.731	0.237	0.475
316	M8	a6	none	xgbTree	42	146	35	1048	0.858	0.712	0.317	0.223	0.545	0.685	0.266	0.514
317	M8	a6	none	xgbLinear	62	126	110	973	0.814	0.623	0.344	0.330	0.360	0.731	0.262	0.484
318	M8	a6	smote	logistic_reg	61	127	315	768	0.652	0.510	0.216	0.324	0.162	0.567	0.415	0.363
319	M8	a6	smote	decision_tree_gini	87	101	329	754	0.662	0.546	0.288	0.463	0.209	0.622	0.395	0.417

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
320	M8	a6	smote	decision_tree_entropy	85	103	283	800	0.696	0.558	0.306	0.452	0.231	0.592	0.417	0.432
321	M8	a6	smote	random_forest	156	32	820	263	0.330	0.526	0.268	0.830	0.160	0.617	0.598	0.397
322	M8	a6	smote	xgbTree	155	33	725	358	0.404	0.546	0.290	0.824	0.176	0.680	0.546	0.418
323	M8	a6	smote	xgbLinear	148	40	719	364	0.403	0.536	0.281	0.787	0.171	0.686	0.572	0.408
324	M8	a6	rose	logistic_reg	74	114	337	746	0.645	0.524	0.247	0.394	0.180	0.576	0.464	0.385
325	M8	a6	rose	decision_tree_gini	25	163	23	1060	0.854	0.694	0.212	0.133	0.521	0.432	0.391	0.453
326	M8	a6	rose	decision_tree_entropy	44	144	91	992	0.815	0.600	0.272	0.234	0.326	0.380	0.382	0.436
327	M8	a6	rose	random_forest	48	140	96	987	0.814	0.605	0.289	0.255	0.333	0.619	0.397	0.447
328	M8	a6	rose	xgbTree	18	170	88	995	0.797	0.512	0.122	0.096	0.170	0.575	0.380	0.317
329	M8	a6	rose	xgbLinear	56	132	134	949	0.791	0.586	0.296	0.298	0.295	0.610	0.341	0.441
330	M8	a9	none	logistic_reg	0	188	0	1083	0.852			0.000		0.590	0.285	
331	M8	a9	none	decision_tree_gini	19	169	37	1046	0.838	0.600	0.156	0.101	0.339	0.630	0.275	0.378
332	M8	a9	none	decision_tree_entropy	35	153	149	934	0.762	0.525	0.188	0.186	0.190	0.597	0.297	0.356
333	M8	a9	none	random_forest	14	174	15	1068	0.851	0.671	0.129	0.074	0.483	0.676	0.236	0.400
334	M8	a9	none	xgbTree	32	156	60	1023	0.830	0.608	0.229	0.170	0.348	0.731	0.265	0.418
335	M8	a9	none	xgbLinear	38	150	52	1031	0.841	0.648	0.273	0.202	0.422	0.702	0.236	0.460
336	M8	a9	smote	logistic_reg	60	128	213	870	0.732	0.546	0.260	0.319	0.220	0.593	0.403	0.403
337	M8	a9	smote	decision_tree_gini	98	90	286	797	0.704	0.577	0.343	0.521	0.255	0.667	0.387	0.460
338	M8	a9	smote	decision_tree_entropy	98	90	286	797	0.704	0.577	0.343	0.521	0.255	0.666	0.389	0.460
339	M8	a9	smote	random_forest	148	40	716	367	0.405	0.537	0.281	0.787	0.171	0.652	0.557	0.409
340	M8	a9	smote	xgbTree	137	51	446	637	0.609	0.580	0.355	0.729	0.235	0.722	0.443	0.468
341	M8	a9	smote	xgbLinear	157	31	725	358	0.405	0.549	0.293	0.835	0.178	0.686	0.559	0.421
342	M8	a9	rose	logistic_reg	71	117	319	764	0.657	0.525	0.246	0.378	0.182	0.585	0.459	0.385
343	M8	a9	rose	decision_tree_gini	73	115	183	900	0.766	0.586	0.329	0.388	0.285	0.429	0.416	0.457
344	M8	a9	rose	decision_tree_entropy	9	179	7	1076	0.854	0.710	0.088	0.048	0.562	0.443	0.395	0.399
345	M8	a9	rose	random_forest	52	136	164	919	0.764	0.556	0.257	0.277	0.241	0.585	0.404	0.407
346	M8	a9	rose	xgbTree	51	137	286	797	0.667	0.502	0.194	0.271	0.151	0.509	0.401	0.348

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
347	M8	a9	rose	xgbLinear	26	162	207	876	0.710	0.478	0.124	0.138	0.112	0.518	0.362	0.301
348	M8	all	none	logistic_reg	0	188	0	1083	0.852			0.000		0.561	0.287	
349	M8	all	none	decision_tree_gini	63	125	81	1002	0.838	0.663	0.380	0.335	0.438	0.661	0.278	0.521
350	M8	all	none	decision_tree_entropy	40	148	40	1043	0.852	0.688	0.299	0.213	0.500	0.659	0.270	0.493
351	M8	all	none	random_forest	21	167	11	1072	0.860	0.761	0.191	0.112	0.656	0.691	0.259	0.476
352	M8	all	none	xgbTree	37	151	41	1042	0.849	0.674	0.278	0.197	0.474	0.698	0.246	0.476
353	M8	all	none	xgbLinear	43	145	61	1022	0.838	0.645	0.295	0.229	0.413	0.694	0.222	0.470
354	M8	all	smote	logistic_reg	57	131	276	807	0.680	0.516	0.219	0.303	0.171	0.594	0.397	0.367
355	M8	all	smote	decision_tree_gini	105	83	327	756	0.677	0.572	0.339	0.559	0.243	0.664	0.382	0.455
356	M8	all	smote	decision_tree_entropy	106	82	294	789	0.704	0.585	0.361	0.564	0.265	0.667	0.403	0.473
357	M8	all	smote	random_forest	167	21	977	106	0.215	0.490	0.251	0.888	0.146	0.660	0.601	0.371
358	M8	all	smote	xgbTree	129	59	566	517	0.508	0.542	0.292	0.686	0.186	0.636	0.502	0.417
359	M8	all	smote	xgbLinear	151	37	697	386	0.423	0.545	0.292	0.803	0.178	0.698	0.570	0.418
360	M8	all	rose	logistic_reg	68	120	341	742	0.637	0.514	0.228	0.362	0.166	0.555	0.466	0.371
361	M8	all	rose	decision_tree_gini	72	116	152	931	0.789	0.605	0.350	0.383	0.321	0.402	0.429	0.477
362	M8	all	rose	decision_tree_entropy	72	116	152	931	0.789	0.605	0.350	0.383	0.321	0.402	0.430	0.477
363	M8	all	rose	random_forest	76	112	276	807	0.695	0.547	0.281	0.404	0.216	0.605	0.437	0.414
364	M8	all	rose	xgbTree	49	139	137	946	0.783	0.568	0.262	0.261	0.263	0.553	0.425	0.415
365	M8	all	rose	xgbLinear	73	115	261	822	0.704	0.548	0.280	0.388	0.219	0.573	0.410	0.414
366	M9	a0	none	logistic_reg	0	188	0	1083	0.852			0.000		0.612	0.284	
367	M9	a0	none	decision_tree_gini	51	137	113	970	0.803	0.594	0.290	0.271	0.311	0.653	0.280	0.442
368	M9	a0	none	decision_tree_entropy	41	147	68	1015	0.831	0.625	0.276	0.218	0.376	0.655	0.276	0.450
369	M9	a0	none	random_forest	9	179	6	1077	0.854	0.729	0.089	0.048	0.600	0.656	0.276	0.409
370	M9	a0	none	xgbTree	45	143	100	983	0.809	0.592	0.270	0.239	0.310	0.642	0.285	0.431
371	M9	a0	none	xgbLinear	47	141	79	1004	0.827	0.625	0.299	0.250	0.373	0.704	0.252	0.462
372	M9	a0	smote	logistic_reg	81	107	293	790	0.685	0.549	0.288	0.431	0.217	0.615	0.394	0.418
373	M9	a0	smote	decision_tree_gini	72	116	296	787	0.676	0.534	0.259	0.383	0.196	0.610	0.403	0.396

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
374	M9	a0	smote	decision_tree_entropy	119	69	361	722	0.662	0.580	0.356	0.633	0.248	0.617	0.416	0.468
375	M9	a0	smote	random_forest	168	20	833	250	0.329	0.547	0.283	0.894	0.168	0.595	0.605	0.415
376	M9	a0	smote	xgbTree	147	41	692	391	0.423	0.540	0.286	0.782	0.175	0.643	0.550	0.413
377	M9	a0	smote	xgbLinear	148	40	706	377	0.413	0.539	0.284	0.787	0.173	0.614	0.570	0.411
378	M9	a0	rose	logistic_reg	59	129	282	801	0.677	0.517	0.223	0.314	0.173	0.582	0.457	0.370
379	M9	a0	rose	decision_tree_gini	115	73	458	625	0.582	0.548	0.302	0.612	0.201	0.597	0.463	0.425
380	M9	a0	rose	decision_tree_entropy	112	76	475	608	0.566	0.540	0.289	0.596	0.191	0.611	0.486	0.414
381	M9	a0	rose	random_forest	105	83	490	593	0.549	0.527	0.268	0.559	0.176	0.600	0.473	0.398
382	M9	a0	rose	xgbTree	130	58	624	459	0.463	0.530	0.276	0.691	0.172	0.571	0.510	0.403
383	M9	a0	rose	xgbLinear	139	49	676	407	0.430	0.532	0.277	0.739	0.171	0.612	0.533	0.404
384	M9	a3	none	logistic_reg	0	188	0	1083	0.852			0.000		0.568	0.289	
385	M9	a3	none	decision_tree_gini	50	138	111	972	0.804	0.593	0.287	0.266	0.311	0.639	0.283	0.440
386	M9	a3	none	decision_tree_entropy	41	147	88	995	0.815	0.595	0.259	0.218	0.318	0.656	0.278	0.427
387	M9	a3	none	random_forest	14	174	14	1069	0.852	0.680	0.130	0.074	0.500	0.682	0.261	0.405
388	M9	a3	none	xgbTree	20	168	77	1006	0.807	0.532	0.140	0.106	0.206	0.644	0.267	0.336
389	M9	a3	none	xgbLinear	35	153	110	973	0.793	0.553	0.210	0.186	0.241	0.652	0.264	0.381
390	M9	a3	smote	logistic_reg	72	116	272	811	0.695	0.542	0.271	0.383	0.209	0.591	0.411	0.406
391	M9	a3	smote	decision_tree_gini	106	82	422	661	0.603	0.545	0.296	0.564	0.201	0.640	0.412	0.421
392	M9	a3	smote	decision_tree_entropy	118	70	412	671	0.621	0.564	0.329	0.628	0.223	0.637	0.416	0.446
393	M9	a3	smote	random_forest	167	21	813	270	0.344	0.549	0.286	0.888	0.170	0.609	0.611	0.418
394	M9	a3	smote	xgbTree	148	40	667	416	0.444	0.547	0.295	0.787	0.182	0.631	0.541	0.421
395	M9	a3	smote	xgbLinear	153	35	681	402	0.437	0.552	0.299	0.814	0.183	0.627	0.547	0.426
396	M9	a3	rose	logistic_reg	63	125	314	769	0.655	0.514	0.223	0.335	0.167	0.548	0.462	0.368
397	M9	a3	rose	decision_tree_gini	92	96	513	570	0.521	0.504	0.232	0.489	0.152	0.566	0.486	0.368
398	M9	a3	rose	decision_tree_entropy	119	69	661	422	0.426	0.506	0.246	0.633	0.153	0.422	0.485	0.376
399	M9	a3	rose	random_forest	99	89	437	646	0.586	0.532	0.273	0.527	0.185	0.605	0.453	0.403
400	M9	a3	rose	xgbTree	108	80	505	578	0.540	0.527	0.270	0.574	0.176	0.598	0.474	0.398

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
401	M9	a3	rose	xgbLinear	123	65	605	478	0.473	0.525	0.269	0.654	0.169	0.580	0.509	0.397
402	M9	a6	none	logistic_reg	0	188	0	1083	0.852			0.000		0.595	0.287	
403	M9	a6	none	decision_tree_gini	48	140	80	1003	0.827	0.626	0.304	0.255	0.375	0.657	0.283	0.465
404	M9	a6	none	decision_tree_entropy	38	150	58	1025	0.836	0.634	0.268	0.202	0.396	0.662	0.276	0.451
405	M9	a6	none	random_forest	23	165	13	1070	0.860	0.753	0.205	0.122	0.639	0.724	0.255	0.479
406	M9	a6	none	xgbTree	51	137	58	1025	0.847	0.675	0.343	0.271	0.468	0.719	0.272	0.509
407	M9	a6	none	xgbLinear	55	133	99	984	0.817	0.619	0.322	0.293	0.357	0.696	0.267	0.470
408	M9	a6	smote	logistic_reg	79	109	317	766	0.665	0.537	0.271	0.420	0.199	0.590	0.413	0.404
409	M9	a6	smote	decision_tree_gini	74	114	286	797	0.685	0.540	0.270	0.394	0.206	0.623	0.389	0.405
410	M9	a6	smote	decision_tree_entropy	85	103	299	784	0.684	0.553	0.297	0.452	0.221	0.591	0.418	0.425
411	M9	a6	smote	random_forest	156	32	742	341	0.391	0.544	0.287	0.830	0.174	0.609	0.584	0.416
412	M9	a6	smote	xgbTree	149	39	619	464	0.482	0.558	0.312	0.793	0.194	0.680	0.507	0.435
413	M9	a6	smote	xgbLinear	151	37	704	379	0.417	0.544	0.290	0.803	0.177	0.624	0.574	0.417
414	M9	a6	rose	logistic_reg	78	110	335	748	0.650	0.530	0.260	0.415	0.189	0.573	0.460	0.395
415	M9	a6	rose	decision_tree_gini	101	87	483	600	0.552	0.523	0.262	0.537	0.173	0.572	0.470	0.392
416	M9	a6	rose	decision_tree_entropy	106	82	471	612	0.565	0.533	0.277	0.564	0.184	0.593	0.473	0.405
417	M9	a6	rose	random_forest	105	83	411	672	0.611	0.547	0.298	0.559	0.203	0.633	0.452	0.423
418	M9	a6	rose	xgbTree	113	75	489	594	0.556	0.538	0.286	0.601	0.188	0.620	0.468	0.412
419	M9	a6	rose	xgbLinear	118	70	538	545	0.522	0.533	0.280	0.628	0.180	0.604	0.481	0.406
420	M9	a9	none	logistic_reg	0	188	0	1083	0.852			0.000		0.647	0.278	
421	M9	a9	none	decision_tree_gini	19	169	37	1046	0.838	0.600	0.156	0.101	0.339	0.630	0.276	0.378
422	M9	a9	none	decision_tree_entropy	37	151	131	952	0.778	0.542	0.208	0.197	0.220	0.670	0.291	0.375
423	M9	a9	none	random_forest	16	172	20	1063	0.849	0.653	0.143	0.085	0.444	0.692	0.256	0.398
424	M9	a9	none	xgbTree	32	156	61	1022	0.829	0.606	0.228	0.170	0.344	0.703	0.259	0.417
425	M9	a9	none	xgbLinear	38	150	64	1019	0.832	0.622	0.262	0.202	0.373	0.698	0.250	0.442
426	M9	a9	smote	logistic_reg	74	114	204	879	0.750	0.576	0.318	0.394	0.266	0.639	0.393	0.447
427	M9	a9	smote	decision_tree_gini	100	88	295	788	0.699	0.576	0.343	0.532	0.253	0.645	0.375	0.460

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
428	M9	a9	smote	decision_tree_entropy	98	90	299	784	0.694	0.572	0.335	0.521	0.247	0.658	0.394	0.453
429	M9	a9	smote	random_forest	142	46	642	441	0.459	0.543	0.292	0.755	0.181	0.639	0.549	0.418
430	M9	a9	smote	xgbTree	141	47	621	462	0.474	0.546	0.297	0.750	0.185	0.676	0.510	0.422
431	M9	a9	smote	xgbLinear	158	30	749	334	0.387	0.546	0.289	0.840	0.174	0.673	0.581	0.417
432	M9	a9	rose	logistic_reg	76	112	317	766	0.662	0.533	0.262	0.404	0.193	0.604	0.452	0.397
433	M9	a9	rose	decision_tree_gini	108	80	424	659	0.603	0.547	0.300	0.574	0.203	0.618	0.465	0.424
434	M9	a9	rose	decision_tree_entropy	138	50	579	504	0.505	0.551	0.305	0.734	0.192	0.366	0.461	0.428
435	M9	a9	rose	random_forest	105	83	470	613	0.565	0.532	0.275	0.559	0.183	0.612	0.464	0.403
436	M9	a9	rose	xgbTree	136	52	592	491	0.493	0.546	0.297	0.723	0.187	0.616	0.475	0.421
437	M9	a9	rose	xgbLinear	137	51	675	408	0.429	0.529	0.274	0.729	0.169	0.604	0.498	0.401
438	M9	all	none	logistic_reg	0	188	0	1083	0.852			0.000		0.592	0.285	
439	M9	all	none	decision_tree_gini	62	126	80	1003	0.838	0.663	0.376	0.330	0.437	0.657	0.277	0.519
440	M9	all	none	decision_tree_entropy	39	149	39	1044	0.852	0.688	0.293	0.207	0.500	0.655	0.270	0.490
441	M9	all	none	random_forest	28	160	18	1065	0.860	0.739	0.239	0.149	0.609	0.681	0.277	0.489
442	M9	all	none	xgbTree	44	144	68	1015	0.833	0.634	0.293	0.234	0.393	0.687	0.273	0.464
443	M9	all	none	xgbLinear	45	143	53	1030	0.846	0.669	0.315	0.239	0.459	0.727	0.228	0.492
444	M9	all	smote	logistic_reg	67	121	246	837	0.711	0.544	0.267	0.356	0.214	0.621	0.393	0.406
445	M9	all	smote	decision_tree_gini	102	86	282	801	0.710	0.584	0.357	0.543	0.266	0.664	0.366	0.470
446	M9	all	smote	decision_tree_entropy	118	70	321	762	0.692	0.592	0.376	0.628	0.269	0.674	0.397	0.484
447	M9	all	smote	random_forest	159	29	794	289	0.352	0.538	0.279	0.846	0.167	0.648	0.582	0.408
448	M9	all	smote	xgbTree	147	41	630	453	0.472	0.553	0.305	0.782	0.189	0.700	0.513	0.429
449	M9	all	smote	xgbLinear	143	45	619	464	0.478	0.550	0.301	0.761	0.188	0.684	0.512	0.425
450	M9	all	rose	logistic_reg	70	118	363	720	0.622	0.510	0.225	0.372	0.162	0.550	0.468	0.368
451	M9	all	rose	decision_tree_gini	124	64	707	376	0.393	0.502	0.243	0.660	0.149	0.548	0.525	0.373
452	M9	all	rose	decision_tree_entropy	139	49	688	395	0.420	0.529	0.274	0.739	0.168	0.412	0.511	0.401
453	M9	all	rose	random_forest	107	81	425	658	0.602	0.546	0.297	0.569	0.201	0.643	0.465	0.421
454	M9	all	rose	xgbTree	170	18	807	276	0.351	0.556	0.292	0.904	0.174	0.667	0.515	0.424

Table A.3: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
455	M9	all	rose	xgbLinear	112	76	501	582	0.546	0.534	0.280	0.596	0.183	0.593	0.448	0.407

A.2.2 Object Recognition

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
1	M4	a0	none	rule-based	88	200	100	883	0.764	0.642	0.370	0.306	0.468	0.670	0.236	0.506
2	M4	a3	none	rule-based	77	227	111	856	0.734	0.600	0.313	0.253	0.410	0.670	0.266	0.456
3	M4	a6	none	rule-based	58	166	130	917	0.767	0.578	0.282	0.259	0.309	0.670	0.233	0.430
4	M4	a9	none	rule-based	26	118	162	965	0.780	0.515	0.157	0.181	0.138	0.670	0.220	0.336
5	M4	all	none	rule-based	97	318	91	765	0.678	0.611	0.322	0.234	0.516	0.670	0.322	0.466
6	M5	a0	none	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.786	
7	M5	a0	none	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
8	M5	a0	none	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
9	M5	a0	none	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.500	1.000	
10	M5	a0	none	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.788	
11	M5	a0	none	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.786	
12	M5	a0	smote	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.563	
13	M5	a0	smote	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.571	
14	M5	a0	smote	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.571	
15	M5	a0	smote	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.866	
16	M5	a0	smote	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.565	
17	M5	a0	smote	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.563	
18	M5	a0	rose	logistic_reg	86	58	14	30	0.617	0.600	0.705	0.597	0.860	0.640	0.482	0.653
19	M5	a0	rose	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.554	
20	M5	a0	rose	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.558	
21	M5	a0	rose	random_forest	86	58	14	30	0.617	0.600	0.705	0.597	0.860	0.640	0.673	0.653
22	M5	a0	rose	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.561	
23	M5	a0	rose	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.640	0.727	
24	M5	a3	none	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.633	0.788	
25	M5	a3	none	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
26	M5	a3	none	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
27	M5	a3	none	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.500	1.000	
28	M5	a3	none	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.633	0.790	
29	M5	a3	none	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.633	0.788	
30	M5	a3	smote	logistic_reg	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.568	0.668
31	M5	a3	smote	decision_tree_gini	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.568	0.668
32	M5	a3	smote	decision_tree_entropy	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.568	0.668
33	M5	a3	smote	random_forest	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.783	0.668
34	M5	a3	smote	xgbTree	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.572	0.668
35	M5	a3	smote	xgbLinear	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.568	0.668
36	M5	a3	rose	logistic_reg	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.487	0.668
37	M5	a3	rose	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.633	0.576	
38	M5	a3	rose	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.633	0.576	
39	M5	a3	rose	random_forest	94	50	17	27	0.644	0.599	0.737	0.653	0.847	0.633	0.720	0.668
40	M5	a3	rose	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.633	0.569	
41	M5	a3	rose	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.633	0.763	
42	M5	a6	none	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.390	0.789	
43	M5	a6	none	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
44	M5	a6	none	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
45	M5	a6	none	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.500	1.000	
46	M5	a6	none	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.390	0.792	
47	M5	a6	none	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.390	0.789	
48	M5	a6	smote	logistic_reg	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.573	0.687
49	M5	a6	smote	decision_tree_gini	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.573	0.687
50	M5	a6	smote	decision_tree_entropy	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.573	0.687
51	M5	a6	smote	random_forest	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.692	0.687
52	M5	a6	smote	xgbTree	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.570	0.687

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
53	M5	a6	smote	xgbLinear	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.573	0.687
54	M5	a6	rose	logistic_reg	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.490	0.687
55	M5	a6	rose	decision_tree_gini	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.564	0.687
56	M5	a6	rose	decision_tree_entropy	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.564	0.687
57	M5	a6	rose	random_forest	0	144	0	44	0.234			0.000		0.390	0.296	
58	M5	a6	rose	xgbTree	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.608	0.687
59	M5	a6	rose	xgbLinear	107	37	23	21	0.681	0.593	0.781	0.743	0.823	0.390	0.492	0.687
60	M5	a9	none	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.516	0.790	
61	M5	a9	none	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
62	M5	a9	none	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
63	M5	a9	none	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.500	1.000	
64	M5	a9	none	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.516	0.793	
65	M5	a9	none	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.516	0.790	
66	M5	a9	smote	logistic_reg	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.574	0.640
67	M5	a9	smote	decision_tree_gini	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.574	0.640
68	M5	a9	smote	decision_tree_entropy	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.574	0.640
69	M5	a9	smote	random_forest	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.870	0.640
70	M5	a9	smote	xgbTree	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.570	0.640
71	M5	a9	smote	xgbLinear	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.574	0.640
72	M5	a9	rose	logistic_reg	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.490	0.640
73	M5	a9	rose	decision_tree_gini	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.575	0.640
74	M5	a9	rose	decision_tree_entropy	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.575	0.640
75	M5	a9	rose	random_forest	0	144	0	44	0.234			0.000		0.516	0.345	
76	M5	a9	rose	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.516	0.626	
77	M5	a9	rose	xgbLinear	123	21	39	5	0.681	0.476	0.804	0.854	0.759	0.516	0.683	0.640
78	M5	all	none	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.791	
79	M5	all	none	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
80	M5	all	none	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.790	
81	M5	all	none	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.500	1.000	
82	M5	all	none	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.794	
83	M5	all	none	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.791	
84	M5	all	smote	logistic_reg	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.575	
85	M5	all	smote	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.571	
86	M5	all	smote	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.500	0.571	
87	M5	all	smote	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.898	
88	M5	all	smote	xgbTree	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.571	
89	M5	all	smote	xgbLinear	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.575	
90	M5	all	rose	logistic_reg	80	64	11	33	0.601	0.610	0.681	0.556	0.879	0.653	0.492	0.645
91	M5	all	rose	decision_tree_gini	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.569	
92	M5	all	rose	decision_tree_entropy	144	0	44	0	0.766		0.867	1.000	0.766	0.653	0.569	
93	M5	all	rose	random_forest	80	64	11	33	0.601	0.610	0.681	0.556	0.879	0.653	0.571	0.645
94	M5	all	rose	xgbTree	80	64	11	33	0.601	0.610	0.681	0.556	0.879	0.653	0.563	0.645
95	M5	all	rose	xgbLinear	80	64	11	33	0.601	0.610	0.681	0.556	0.879	0.653	0.616	0.645
96	M6	a0	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.633	0.898	0.878
97	M6	a0	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
98	M6	a0	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
99	M6	a0	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.610	0.866	0.878
100	M6	a0	none	xgbTree	130	14	33	11	0.750	0.619	0.847	0.903	0.798	0.602	0.845	0.733
101	M6	a0	none	xgbLinear	131	13	32	12	0.761	0.642	0.853	0.910	0.804	0.577	0.840	0.748
102	M6	a0	smote	logistic_reg	136	8	41	3	0.739	0.521	0.847	0.944	0.768	0.672	0.735	0.684
103	M6	a0	smote	decision_tree_gini	120	24	33	11	0.697	0.549	0.808	0.833	0.784	0.589	0.665	0.679
104	M6	a0	smote	decision_tree_entropy	103	41	24	20	0.654	0.569	0.760	0.715	0.811	0.489	0.595	0.665
105	M6	a0	smote	random_forest	72	72	16	28	0.532	0.549	0.621	0.500	0.818	0.648	0.480	0.585
106	M6	a0	smote	xgbTree	76	68	15	29	0.559	0.567	0.647	0.528	0.835	0.622	0.497	0.607

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
107	M6	a0	smote	xgbLinear	76	68	12	32	0.574	0.592	0.655	0.528	0.864	0.669	0.470	0.623
108	M6	a0	rose	logistic_reg	128	16	34	10	0.734	0.587	0.837	0.889	0.790	0.636	0.688	0.712
109	M6	a0	rose	decision_tree_gini	126	18	41	3	0.686	0.449	0.810	0.875	0.754	0.671	0.733	0.629
110	M6	a0	rose	decision_tree_entropy	126	18	41	3	0.686	0.449	0.810	0.875	0.754	0.671	0.733	0.629
111	M6	a0	rose	random_forest	112	32	32	12	0.660	0.525	0.778	0.778	0.778	0.634	0.657	0.652
112	M6	a0	rose	xgbTree	126	18	24	20	0.777	0.683	0.857	0.875	0.840	0.689	0.730	0.770
113	M6	a0	rose	xgbLinear	124	20	32	12	0.723	0.585	0.827	0.861	0.795	0.611	0.743	0.706
114	M6	a3	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.651	0.897	0.878
115	M6	a3	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
116	M6	a3	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
117	M6	a3	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.419	0.863	0.878
118	M6	a3	none	xgbTree	140	4	40	4	0.766	0.639	0.864	0.972	0.778	0.590	0.894	0.752
119	M6	a3	none	xgbLinear	133	11	32	12	0.771	0.664	0.861	0.924	0.806	0.590	0.852	0.762
120	M6	a3	smote	logistic_reg	137	7	41	3	0.745	0.535	0.851	0.951	0.770	0.684	0.728	0.693
121	M6	a3	smote	decision_tree_gini	112	32	33	11	0.654	0.514	0.775	0.778	0.772	0.564	0.674	0.645
122	M6	a3	smote	decision_tree_entropy	104	40	29	15	0.633	0.527	0.751	0.722	0.782	0.637	0.595	0.639
123	M6	a3	smote	random_forest	77	67	11	33	0.585	0.603	0.664	0.535	0.875	0.670	0.490	0.633
124	M6	a3	smote	xgbTree	72	72	8	36	0.574	0.617	0.643	0.500	0.900	0.646	0.432	0.630
125	M6	a3	smote	xgbLinear	85	59	12	32	0.622	0.614	0.705	0.590	0.876	0.685	0.517	0.660
126	M6	a3	rose	logistic_reg	131	13	34	10	0.750	0.614	0.848	0.910	0.794	0.640	0.685	0.731
127	M6	a3	rose	decision_tree_gini	116	28	39	5	0.644	0.450	0.776	0.806	0.748	0.674	0.709	0.613
128	M6	a3	rose	decision_tree_entropy	122	22	42	2	0.660	0.414	0.792	0.847	0.744	0.677	0.703	0.603
129	M6	a3	rose	random_forest	117	27	32	12	0.686	0.546	0.799	0.812	0.785	0.603	0.664	0.673
130	M6	a3	rose	xgbTree	129	15	32	12	0.750	0.623	0.846	0.896	0.801	0.663	0.727	0.734
131	M6	a3	rose	xgbLinear	118	26	32	12	0.691	0.551	0.803	0.819	0.787	0.606	0.741	0.677
132	M6	a6	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.654	0.897	0.878
133	M6	a6	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
134	M6	a6	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
135	M6	a6	none	random_forest	141	3	39	5	0.777	0.704	0.870	0.979	0.783	0.596	0.822	0.787
136	M6	a6	none	xgbTree	139	5	37	7	0.777	0.687	0.869	0.965	0.790	0.631	0.870	0.778
137	M6	a6	none	xgbLinear	132	12	34	10	0.755	0.625	0.852	0.917	0.795	0.631	0.868	0.738
138	M6	a6	smote	logistic_reg	135	9	32	12	0.782	0.690	0.868	0.938	0.808	0.710	0.727	0.779
139	M6	a6	smote	decision_tree_gini	100	44	20	24	0.660	0.593	0.758	0.694	0.833	0.656	0.567	0.675
140	M6	a6	smote	decision_tree_entropy	99	45	19	25	0.660	0.598	0.756	0.688	0.839	0.631	0.562	0.677
141	M6	a6	smote	random_forest	61	83	16	28	0.473	0.522	0.552	0.424	0.792	0.617	0.468	0.537
142	M6	a6	smote	xgbTree	69	75	7	37	0.564	0.619	0.627	0.479	0.908	0.727	0.437	0.623
143	M6	a6	smote	xgbLinear	78	66	9	35	0.601	0.622	0.675	0.542	0.897	0.662	0.470	0.648
144	M6	a6	rose	logistic_reg	123	21	31	13	0.723	0.591	0.826	0.854	0.799	0.661	0.678	0.708
145	M6	a6	rose	decision_tree_gini	121	23	35	9	0.691	0.528	0.807	0.840	0.776	0.580	0.704	0.668
146	M6	a6	rose	decision_tree_entropy	121	23	35	9	0.691	0.528	0.807	0.840	0.776	0.580	0.713	0.668
147	M6	a6	rose	random_forest	119	25	33	11	0.691	0.544	0.804	0.826	0.783	0.621	0.652	0.674
148	M6	a6	rose	xgbTree	117	27	32	12	0.686	0.546	0.799	0.812	0.785	0.623	0.722	0.673
149	M6	a6	rose	xgbLinear	124	20	34	10	0.713	0.559	0.821	0.861	0.785	0.645	0.736	0.690
150	M6	a9	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.672	0.895	0.878
151	M6	a9	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
152	M6	a9	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
153	M6	a9	none	random_forest	141	3	37	7	0.787	0.746	0.876	0.979	0.792	0.614	0.820	0.811
154	M6	a9	none	xgbTree	136	8	35	9	0.771	0.662	0.863	0.944	0.795	0.619	0.858	0.763
155	M6	a9	none	xgbLinear	129	15	31	13	0.755	0.635	0.849	0.896	0.806	0.573	0.858	0.742
156	M6	a9	smote	logistic_reg	131	13	30	14	0.771	0.666	0.859	0.910	0.814	0.725	0.713	0.763
157	M6	a9	smote	decision_tree_gini	88	56	22	22	0.585	0.541	0.693	0.611	0.800	0.537	0.533	0.617
158	M6	a9	smote	decision_tree_entropy	83	61	8	36	0.633	0.642	0.706	0.576	0.912	0.675	0.427	0.674
159	M6	a9	smote	random_forest	75	69	8	36	0.590	0.623	0.661	0.521	0.904	0.693	0.452	0.642
160	M6	a9	smote	xgbTree	64	80	18	26	0.479	0.513	0.566	0.444	0.780	0.596	0.420	0.540

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
161	M6	a9	smote	xgbLinear	71	73	12	32	0.548	0.580	0.626	0.493	0.855	0.641	0.465	0.603
162	M6	a9	rose	logistic_reg	127	17	27	17	0.766	0.662	0.852	0.882	0.825	0.675	0.676	0.757
163	M6	a9	rose	decision_tree_gini	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.687	0.639
164	M6	a9	rose	decision_tree_entropy	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.687	0.639
165	M6	a9	rose	random_forest	120	24	31	13	0.707	0.573	0.814	0.833	0.795	0.600	0.663	0.693
166	M6	a9	rose	xgbTree	112	32	32	12	0.660	0.525	0.778	0.778	0.778	0.628	0.700	0.652
167	M6	a9	rose	xgbLinear	113	31	32	12	0.665	0.529	0.782	0.785	0.779	0.603	0.738	0.656
168	M6	all	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.657	0.896	0.878
169	M6	all	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
170	M6	all	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
171	M6	all	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.594	0.871	0.878
172	M6	all	none	xgbTree	140	4	38	6	0.777	0.693	0.870	0.972	0.787	0.635	0.886	0.781
173	M6	all	none	xgbLinear	138	6	36	8	0.777	0.682	0.868	0.958	0.793	0.590	0.883	0.775
174	M6	all	smote	logistic_reg	132	12	36	8	0.745	0.593	0.846	0.917	0.786	0.705	0.716	0.720
175	M6	all	smote	decision_tree_gini	106	38	27	17	0.654	0.553	0.765	0.736	0.797	0.583	0.646	0.659
176	M6	all	smote	decision_tree_entropy	106	38	27	17	0.654	0.553	0.765	0.736	0.797	0.533	0.649	0.659
177	M6	all	smote	random_forest	74	70	21	23	0.516	0.513	0.619	0.514	0.779	0.626	0.470	0.566
178	M6	all	smote	xgbTree	60	84	9	35	0.505	0.582	0.563	0.417	0.870	0.642	0.408	0.573
179	M6	all	smote	xgbLinear	60	84	13	31	0.484	0.546	0.553	0.417	0.822	0.643	0.394	0.549
180	M6	all	rose	logistic_reg	122	22	27	17	0.739	0.627	0.833	0.847	0.819	0.665	0.677	0.730
181	M6	all	rose	decision_tree_gini	127	17	42	2	0.686	0.428	0.812	0.882	0.751	0.442	0.737	0.620
182	M6	all	rose	decision_tree_entropy	127	17	42	2	0.686	0.428	0.812	0.882	0.751	0.442	0.737	0.620
183	M6	all	rose	random_forest	115	29	32	12	0.676	0.537	0.790	0.799	0.782	0.600	0.646	0.664
184	M6	all	rose	xgbTree	129	15	33	11	0.745	0.610	0.843	0.896	0.796	0.651	0.734	0.726
185	M6	all	rose	xgbLinear	116	28	23	21	0.729	0.632	0.820	0.806	0.835	0.642	0.708	0.726
186	M7	a0	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.664	0.896	0.878
187	M7	a0	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
188	M7	a0	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
189	M7	a0	none	random_forest	141	3	39	5	0.777	0.704	0.870	0.979	0.783	0.621	0.800	0.787
190	M7	a0	none	xgbTree	136	8	34	10	0.777	0.678	0.866	0.944	0.800	0.602	0.866	0.772
191	M7	a0	none	xgbLinear	132	12	34	10	0.755	0.625	0.852	0.917	0.795	0.631	0.855	0.738
192	M7	a0	smote	logistic_reg	135	9	32	12	0.782	0.690	0.868	0.938	0.808	0.714	0.722	0.779
193	M7	a0	smote	decision_tree_gini	65	79	9	35	0.532	0.593	0.596	0.451	0.878	0.626	0.416	0.595
194	M7	a0	smote	decision_tree_entropy	88	56	17	27	0.612	0.582	0.707	0.611	0.838	0.621	0.533	0.644
195	M7	a0	smote	random_forest	84	60	16	28	0.596	0.579	0.689	0.583	0.840	0.635	0.515	0.634
196	M7	a0	smote	xgbTree	74	70	8	36	0.585	0.621	0.655	0.514	0.902	0.691	0.472	0.638
197	M7	a0	smote	xgbLinear	74	70	15	29	0.548	0.562	0.635	0.514	0.831	0.634	0.476	0.599
198	M7	a0	rose	logistic_reg	116	28	25	19	0.718	0.613	0.814	0.806	0.823	0.675	0.676	0.714
199	M7	a0	rose	decision_tree_gini	112	32	23	21	0.707	0.613	0.803	0.778	0.830	0.666	0.650	0.708
200	M7	a0	rose	decision_tree_entropy	112	32	23	21	0.707	0.613	0.803	0.778	0.830	0.666	0.650	0.708
201	M7	a0	rose	random_forest	102	42	23	21	0.654	0.575	0.758	0.708	0.816	0.619	0.604	0.667
202	M7	a0	rose	xgbTree	101	43	24	20	0.644	0.563	0.751	0.701	0.808	0.628	0.593	0.657
203	M7	a0	rose	xgbLinear	110	34	21	23	0.707	0.622	0.800	0.764	0.840	0.692	0.644	0.711
204	M7	a3	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.664	0.896	0.878
205	M7	a3	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
206	M7	a3	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
207	M7	a3	none	random_forest	143	1	39	5	0.787	0.810	0.877	0.993	0.786	0.626	0.805	0.843
208	M7	a3	none	xgbTree	136	8	38	6	0.755	0.605	0.855	0.944	0.782	0.548	0.867	0.730
209	M7	a3	none	xgbLinear	135	9	34	10	0.771	0.663	0.863	0.938	0.799	0.622	0.855	0.763
210	M7	a3	smote	logistic_reg	135	9	33	11	0.777	0.677	0.865	0.938	0.804	0.713	0.722	0.771
211	M7	a3	smote	decision_tree_gini	69	75	9	35	0.553	0.601	0.622	0.479	0.885	0.649	0.430	0.612
212	M7	a3	smote	decision_tree_entropy	69	75	8	36	0.559	0.610	0.624	0.479	0.896	0.651	0.447	0.617
213	M7	a3	smote	random_forest	85	59	17	27	0.596	0.574	0.691	0.590	0.833	0.638	0.514	0.632
214	M7	a3	smote	xgbTree	67	77	12	32	0.527	0.571	0.601	0.465	0.848	0.642	0.465	0.586

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
215	M7	a3	smote	xgbLinear	68	76	14	30	0.521	0.556	0.602	0.472	0.829	0.687	0.442	0.579
216	M7	a3	rose	logistic_reg	116	28	25	19	0.718	0.613	0.814	0.806	0.823	0.675	0.676	0.714
217	M7	a3	rose	decision_tree_gini	112	32	23	21	0.707	0.613	0.803	0.778	0.830	0.666	0.650	0.708
218	M7	a3	rose	decision_tree_entropy	112	32	23	21	0.707	0.613	0.803	0.778	0.830	0.666	0.650	0.708
219	M7	a3	rose	random_forest	101	43	21	23	0.660	0.588	0.759	0.701	0.828	0.648	0.601	0.674
220	M7	a3	rose	xgbTree	101	43	23	21	0.649	0.571	0.754	0.701	0.815	0.631	0.615	0.663
221	M7	a3	rose	xgbLinear	109	35	27	17	0.670	0.564	0.779	0.757	0.801	0.660	0.638	0.671
222	M7	a6	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.664	0.896	0.878
223	M7	a6	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
224	M7	a6	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
225	M7	a6	none	random_forest	137	7	36	8	0.771	0.663	0.864	0.951	0.792	0.620	0.789	0.763
226	M7	a6	none	xgbTree	142	2	40	4	0.777	0.723	0.871	0.986	0.780	0.586	0.887	0.797
227	M7	a6	none	xgbLinear	133	11	35	9	0.755	0.621	0.853	0.924	0.792	0.586	0.873	0.737
228	M7	a6	smote	logistic_reg	135	9	31	13	0.787	0.702	0.871	0.938	0.813	0.720	0.719	0.787
229	M7	a6	smote	decision_tree_gini	93	51	17	27	0.638	0.596	0.732	0.646	0.845	0.607	0.522	0.664
230	M7	a6	smote	decision_tree_entropy	96	48	19	25	0.644	0.589	0.741	0.667	0.835	0.634	0.527	0.665
231	M7	a6	smote	random_forest	82	62	16	28	0.585	0.574	0.678	0.569	0.837	0.624	0.520	0.626
232	M7	a6	smote	xgbTree	81	63	15	29	0.585	0.579	0.675	0.562	0.844	0.668	0.509	0.627
233	M7	a6	smote	xgbLinear	78	66	14	30	0.574	0.580	0.661	0.542	0.848	0.697	0.484	0.621
234	M7	a6	rose	logistic_reg	116	28	25	19	0.718	0.613	0.814	0.806	0.823	0.675	0.677	0.714
235	M7	a6	rose	decision_tree_gini	122	22	26	18	0.745	0.637	0.836	0.847	0.824	0.680	0.677	0.736
236	M7	a6	rose	decision_tree_entropy	122	22	26	18	0.745	0.637	0.836	0.847	0.824	0.680	0.677	0.736
237	M7	a6	rose	random_forest	104	40	25	19	0.654	0.564	0.762	0.722	0.806	0.628	0.620	0.663
238	M7	a6	rose	xgbTree	110	34	24	20	0.691	0.596	0.791	0.764	0.821	0.651	0.629	0.693
239	M7	a6	rose	xgbLinear	104	40	27	17	0.644	0.546	0.756	0.722	0.794	0.608	0.659	0.651
240	M7	a9	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.667	0.896	0.878
241	M7	a9	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
242	M7	a9	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
243	M7	a9	none	random_forest	140	4	37	7	0.782	0.714	0.872	0.972	0.791	0.622	0.806	0.793
244	M7	a9	none	xgbTree	138	6	36	8	0.777	0.682	0.868	0.958	0.793	0.606	0.867	0.775
245	M7	a9	none	xgbLinear	133	11	35	9	0.755	0.621	0.853	0.924	0.792	0.590	0.865	0.737
246	M7	a9	smote	logistic_reg	136	8	32	12	0.787	0.705	0.872	0.944	0.810	0.722	0.721	0.788
247	M7	a9	smote	decision_tree_gini	96	48	17	27	0.654	0.605	0.747	0.667	0.850	0.666	0.537	0.676
248	M7	a9	smote	decision_tree_entropy	96	48	17	27	0.654	0.605	0.747	0.667	0.850	0.636	0.531	0.676
249	M7	a9	smote	random_forest	82	62	16	28	0.585	0.574	0.678	0.569	0.837	0.648	0.531	0.626
250	M7	a9	smote	xgbTree	82	62	14	30	0.596	0.590	0.683	0.569	0.854	0.687	0.507	0.637
251	M7	a9	smote	xgbLinear	79	65	14	30	0.580	0.583	0.667	0.549	0.849	0.662	0.502	0.625
252	M7	a9	rose	logistic_reg	116	28	25	19	0.718	0.613	0.814	0.806	0.823	0.672	0.679	0.714
253	M7	a9	rose	decision_tree_gini	124	20	36	8	0.702	0.530	0.816	0.861	0.775	0.634	0.707	0.673
254	M7	a9	rose	decision_tree_entropy	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.685	0.639
255	M7	a9	rose	random_forest	105	39	22	22	0.676	0.594	0.775	0.729	0.827	0.647	0.611	0.684
256	M7	a9	rose	xgbTree	117	27	24	20	0.729	0.628	0.821	0.812	0.830	0.655	0.671	0.724
257	M7	a9	rose	xgbLinear	103	41	28	16	0.633	0.533	0.749	0.715	0.786	0.618	0.645	0.641
258	M7	all	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.663	0.896	0.878
259	M7	all	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
260	M7	all	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
261	M7	all	none	random_forest	134	10	34	10	0.766	0.649	0.859	0.931	0.798	0.612	0.782	0.754
262	M7	all	none	xgbTree	140	4	38	6	0.777	0.693	0.870	0.972	0.787	0.623	0.876	0.781
263	M7	all	none	xgbLinear	133	11	34	10	0.761	0.636	0.855	0.924	0.796	0.617	0.859	0.746
264	M7	all	smote	logistic_reg	135	9	32	12	0.782	0.690	0.868	0.938	0.808	0.720	0.719	0.779
265	M7	all	smote	decision_tree_gini	93	51	17	27	0.638	0.596	0.732	0.646	0.845	0.607	0.520	0.664
266	M7	all	smote	decision_tree_entropy	93	51	17	27	0.638	0.596	0.732	0.646	0.845	0.607	0.520	0.664
267	M7	all	smote	random_forest	80	64	16	28	0.574	0.569	0.667	0.556	0.833	0.628	0.509	0.618
268	M7	all	smote	xgbTree	71	73	9	35	0.564	0.606	0.634	0.493	0.887	0.679	0.440	0.620

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
269	M7	all	smote	xgbLinear	70	74	15	29	0.527	0.553	0.611	0.486	0.824	0.668	0.453	0.582
270	M7	all	rose	logistic_reg	116	28	25	19	0.718	0.613	0.814	0.806	0.823	0.675	0.676	0.714
271	M7	all	rose	decision_tree_gini	114	30	23	21	0.718	0.622	0.811	0.792	0.832	0.672	0.653	0.717
272	M7	all	rose	decision_tree_entropy	114	30	23	21	0.718	0.622	0.811	0.792	0.832	0.672	0.653	0.717
273	M7	all	rose	random_forest	101	43	23	21	0.649	0.571	0.754	0.701	0.815	0.626	0.609	0.663
274	M7	all	rose	xgbTree	106	38	20	24	0.691	0.614	0.785	0.736	0.841	0.630	0.629	0.700
275	M7	all	rose	xgbLinear	96	48	21	23	0.633	0.572	0.736	0.667	0.821	0.685	0.608	0.654
276	M8	a0	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.685	0.898	0.878
277	M8	a0	none	decision_tree_gini	142	2	40	4	0.777	0.723	0.871	0.986	0.780	0.538	0.910	0.797
278	M8	a0	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
279	M8	a0	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.631	0.857	0.878
280	M8	a0	none	xgbTree	138	6	35	9	0.782	0.699	0.871	0.958	0.798	0.713	0.814	0.785
281	M8	a0	none	xgbLinear	137	7	37	7	0.766	0.644	0.862	0.951	0.787	0.590	0.892	0.753
282	M8	a0	smote	logistic_reg	136	8	34	10	0.777	0.678	0.866	0.944	0.800	0.722	0.729	0.772
283	M8	a0	smote	decision_tree_gini	99	45	16	28	0.676	0.622	0.764	0.688	0.861	0.670	0.496	0.693
284	M8	a0	smote	decision_tree_entropy	96	48	16	28	0.660	0.613	0.750	0.667	0.857	0.670	0.487	0.681
285	M8	a0	smote	random_forest	98	46	23	21	0.633	0.562	0.740	0.681	0.810	0.626	0.576	0.651
286	M8	a0	smote	xgbTree	96	48	24	20	0.617	0.547	0.727	0.667	0.800	0.603	0.614	0.637
287	M8	a0	smote	xgbLinear	88	56	19	25	0.601	0.566	0.701	0.611	0.822	0.647	0.579	0.633
288	M8	a0	rose	logistic_reg	117	27	24	20	0.729	0.628	0.821	0.812	0.830	0.681	0.668	0.724
289	M8	a0	rose	decision_tree_gini	73	71	6	38	0.590	0.636	0.655	0.507	0.924	0.707	0.486	0.646
290	M8	a0	rose	decision_tree_entropy	92	52	15	29	0.644	0.609	0.733	0.639	0.860	0.683	0.582	0.671
291	M8	a0	rose	random_forest	106	38	23	21	0.676	0.589	0.777	0.736	0.822	0.707	0.627	0.683
292	M8	a0	rose	xgbTree	111	33	23	21	0.702	0.609	0.799	0.771	0.828	0.719	0.629	0.704
293	M8	a0	rose	xgbLinear	106	38	24	20	0.670	0.580	0.774	0.736	0.815	0.681	0.665	0.677
294	M8	a3	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.659	0.898	0.878
295	M8	a3	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
296	M8	a3	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
297	M8	a3	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.612	0.859	0.878
298	M8	a3	none	xgbTree	144	0	41	3	0.782	0.889	0.875	1.000	0.778	0.632	0.924	0.882
299	M8	a3	none	xgbLinear	143	1	41	3	0.777	0.764	0.872	0.993	0.777	0.574	0.926	0.818
300	M8	a3	smote	logistic_reg	135	9	31	13	0.787	0.702	0.871	0.938	0.813	0.714	0.721	0.787
301	M8	a3	smote	decision_tree_gini	90	54	31	13	0.548	0.469	0.679	0.625	0.744	0.459	0.566	0.574
302	M8	a3	smote	decision_tree_entropy	101	43	37	7	0.574	0.436	0.716	0.701	0.732	0.547	0.641	0.576
303	M8	a3	smote	random_forest	81	63	23	21	0.543	0.514	0.653	0.562	0.779	0.564	0.548	0.584
304	M8	a3	smote	xgbTree	82	62	20	24	0.564	0.541	0.667	0.569	0.804	0.582	0.527	0.604
305	M8	a3	smote	xgbLinear	81	63	23	21	0.543	0.514	0.653	0.562	0.779	0.570	0.542	0.584
306	M8	a3	rose	logistic_reg	117	27	24	20	0.729	0.628	0.821	0.812	0.830	0.674	0.669	0.724
307	M8	a3	rose	decision_tree_gini	70	74	6	38	0.574	0.630	0.636	0.486	0.921	0.340	0.498	0.633
308	M8	a3	rose	decision_tree_entropy	92	52	15	29	0.644	0.609	0.733	0.639	0.860	0.683	0.580	0.671
309	M8	a3	rose	random_forest	119	25	25	19	0.734	0.629	0.826	0.826	0.826	0.727	0.614	0.728
310	M8	a3	rose	xgbTree	123	21	28	16	0.739	0.624	0.834	0.854	0.815	0.742	0.700	0.729
311	M8	a3	rose	xgbLinear	113	31	31	13	0.670	0.540	0.785	0.785	0.785	0.648	0.683	0.662
312	M8	a6	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.674	0.887	0.878
313	M8	a6	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
314	M8	a6	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
315	M8	a6	none	random_forest	142	2	40	4	0.777	0.723	0.871	0.986	0.780	0.661	0.800	0.797
316	M8	a6	none	xgbTree	143	1	40	4	0.782	0.791	0.875	0.993	0.781	0.662	0.872	0.833
317	M8	a6	none	xgbLinear	135	9	34	10	0.771	0.663	0.863	0.938	0.799	0.584	0.864	0.763
318	M8	a6	smote	logistic_reg	132	12	26	18	0.798	0.718	0.874	0.917	0.835	0.715	0.711	0.796
319	M8	a6	smote	decision_tree_gini	86	58	7	37	0.654	0.657	0.726	0.597	0.925	0.679	0.478	0.691
320	M8	a6	smote	decision_tree_entropy	125	19	34	10	0.718	0.565	0.825	0.868	0.786	0.630	0.722	0.695
321	M8	a6	smote	random_forest	82	62	17	27	0.580	0.566	0.675	0.569	0.828	0.642	0.523	0.620
322	M8	a6	smote	xgbTree	77	67	19	25	0.543	0.537	0.642	0.535	0.802	0.624	0.529	0.589

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
323	M8	a6	smote	xgbLinear	92	52	22	22	0.606	0.552	0.713	0.639	0.807	0.626	0.585	0.633
324	M8	a6	rose	logistic_reg	115	29	24	20	0.718	0.618	0.813	0.799	0.827	0.685	0.651	0.715
325	M8	a6	rose	decision_tree_gini	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.687	0.639
326	M8	a6	rose	decision_tree_entropy	104	40	16	28	0.702	0.639	0.788	0.722	0.867	0.711	0.586	0.714
327	M8	a6	rose	random_forest	107	37	20	24	0.697	0.618	0.790	0.743	0.843	0.708	0.606	0.704
328	M8	a6	rose	xgbTree	106	38	22	22	0.681	0.597	0.779	0.736	0.828	0.670	0.615	0.688
329	M8	a6	rose	xgbLinear	103	41	22	22	0.665	0.587	0.766	0.715	0.824	0.676	0.633	0.676
330	M8	a9	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.680	0.893	0.878
331	M8	a9	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
332	M8	a9	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
333	M8	a9	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.611	0.842	0.878
334	M8	a9	none	xgbTree	141	3	39	5	0.777	0.704	0.870	0.979	0.783	0.700	0.828	0.787
335	M8	a9	none	xgbLinear	134	10	38	6	0.745	0.577	0.848	0.931	0.779	0.674	0.876	0.713
336	M8	a9	smote	logistic_reg	134	10	28	16	0.798	0.721	0.876	0.931	0.827	0.739	0.739	0.799
337	M8	a9	smote	decision_tree_gini	99	45	27	17	0.617	0.530	0.733	0.688	0.786	0.650	0.617	0.632
338	M8	a9	smote	decision_tree_entropy	92	52	25	19	0.590	0.527	0.705	0.639	0.786	0.592	0.575	0.616
339	M8	a9	smote	random_forest	87	57	8	36	0.654	0.651	0.728	0.604	0.916	0.711	0.506	0.690
340	M8	a9	smote	xgbTree	89	55	13	31	0.638	0.617	0.724	0.618	0.873	0.658	0.552	0.670
341	M8	a9	smote	xgbLinear	92	52	9	35	0.676	0.657	0.751	0.639	0.911	0.730	0.524	0.704
342	M8	a9	rose	logistic_reg	118	26	25	19	0.729	0.624	0.822	0.819	0.825	0.682	0.671	0.723
343	M8	a9	rose	decision_tree_gini	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.712	0.639
344	M8	a9	rose	decision_tree_entropy	112	32	37	7	0.633	0.466	0.765	0.778	0.752	0.620	0.679	0.615
345	M8	a9	rose	random_forest	116	28	26	18	0.713	0.604	0.811	0.806	0.817	0.703	0.633	0.708
346	M8	a9	rose	xgbTree	117	27	32	12	0.686	0.546	0.799	0.812	0.785	0.635	0.668	0.673
347	M8	a9	rose	xgbLinear	103	41	27	17	0.638	0.543	0.752	0.715	0.792	0.580	0.656	0.647
348	M8	all	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.707	0.887	0.878
349	M8	all	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
350	M8	all	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
351	M8	all	none	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.694	0.835	
352	M8	all	none	xgbTree	143	1	42	2	0.771	0.720	0.869	0.993	0.773	0.754	0.880	0.795
353	M8	all	none	xgbLinear	143	1	43	1	0.766	0.634	0.867	0.993	0.769	0.756	0.933	0.751
354	M8	all	smote	logistic_reg	137	7	28	16	0.814	0.763	0.887	0.951	0.830	0.736	0.725	0.825
355	M8	all	smote	decision_tree_gini	91	53	28	16	0.569	0.498	0.692	0.632	0.765	0.506	0.613	0.595
356	M8	all	smote	decision_tree_entropy	103	41	34	10	0.601	0.474	0.733	0.715	0.752	0.611	0.624	0.604
357	M8	all	smote	random_forest	104	40	15	29	0.707	0.647	0.791	0.722	0.874	0.678	0.545	0.719
358	M8	all	smote	xgbTree	104	40	16	28	0.702	0.639	0.788	0.722	0.867	0.718	0.640	0.714
359	M8	all	smote	xgbLinear	103	41	18	26	0.686	0.620	0.777	0.715	0.851	0.707	0.626	0.699
360	M8	all	rose	logistic_reg	128	16	28	16	0.766	0.660	0.853	0.889	0.821	0.678	0.683	0.757
361	M8	all	rose	decision_tree_gini	85	59	4	40	0.665	0.680	0.730	0.590	0.955	0.728	0.502	0.705
362	M8	all	rose	decision_tree_entropy	109	35	14	30	0.739	0.674	0.816	0.757	0.886	0.728	0.612	0.745
363	M8	all	rose	random_forest	97	47	11	33	0.691	0.655	0.770	0.674	0.898	0.732	0.563	0.713
364	M8	all	rose	xgbTree	100	44	15	29	0.686	0.633	0.772	0.694	0.870	0.709	0.570	0.703
365	M8	all	rose	xgbLinear	89	55	22	22	0.590	0.544	0.698	0.618	0.802	0.670	0.574	0.621
366	M9	a0	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.683	0.898	0.878
367	M9	a0	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
368	M9	a0	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
369	M9	a0	none	random_forest	143	1	41	3	0.777	0.764	0.872	0.993	0.777	0.653	0.803	0.818
370	M9	a0	none	xgbTree	142	2	40	4	0.777	0.723	0.871	0.986	0.780	0.610	0.873	0.797
371	M9	a0	none	xgbLinear	140	4	38	6	0.777	0.693	0.870	0.972	0.787	0.542	0.900	0.781
372	M9	a0	smote	logistic_reg	135	9	29	15	0.798	0.724	0.877	0.938	0.823	0.727	0.724	0.800
373	M9	a0	smote	decision_tree_gini	115	29	37	7	0.649	0.476	0.777	0.799	0.757	0.671	0.694	0.626
374	M9	a0	smote	decision_tree_entropy	134	10	39	5	0.739	0.554	0.845	0.931	0.775	0.591	0.774	0.700
375	M9	a0	smote	random_forest	103	41	26	18	0.644	0.552	0.755	0.715	0.798	0.609	0.614	0.653
376	M9	a0	smote	xgbTree	112	32	30	14	0.670	0.547	0.783	0.778	0.789	0.620	0.711	0.665

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
377	M9	a0	smote	xgbLinear	99	45	20	24	0.654	0.590	0.753	0.688	0.832	0.644	0.646	0.671
378	M9	a0	rose	logistic_reg	114	30	24	20	0.713	0.613	0.809	0.792	0.826	0.684	0.668	0.711
379	M9	a0	rose	decision_tree_gini	132	12	42	2	0.713	0.451	0.830	0.917	0.759	0.644	0.730	0.640
380	M9	a0	rose	decision_tree_entropy	132	12	42	2	0.713	0.451	0.830	0.917	0.759	0.644	0.715	0.640
381	M9	a0	rose	random_forest	106	38	19	25	0.697	0.622	0.788	0.736	0.848	0.690	0.592	0.705
382	M9	a0	rose	xgbTree	74	70	7	37	0.590	0.630	0.658	0.514	0.914	0.712	0.498	0.644
383	M9	a0	rose	xgbLinear	84	60	10	34	0.628	0.628	0.706	0.583	0.894	0.725	0.522	0.667
384	M9	a3	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.661	0.899	0.878
385	M9	a3	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
386	M9	a3	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
387	M9	a3	none	random_forest	142	2	40	4	0.777	0.723	0.871	0.986	0.780	0.632	0.815	0.797
388	M9	a3	none	xgbTree	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.581	0.904	0.878
389	M9	a3	none	xgbLinear	140	4	41	3	0.761	0.601	0.862	0.972	0.773	0.490	0.915	0.731
390	M9	a3	smote	logistic_reg	134	10	30	14	0.787	0.700	0.870	0.931	0.817	0.714	0.721	0.785
391	M9	a3	smote	decision_tree_gini	115	29	38	6	0.644	0.462	0.774	0.799	0.752	0.546	0.700	0.618
392	M9	a3	smote	decision_tree_entropy	115	29	38	6	0.644	0.462	0.774	0.799	0.752	0.546	0.701	0.618
393	M9	a3	smote	random_forest	89	55	22	22	0.590	0.544	0.698	0.618	0.802	0.591	0.565	0.621
394	M9	a3	smote	xgbTree	95	49	28	16	0.590	0.509	0.712	0.660	0.772	0.580	0.648	0.610
395	M9	a3	smote	xgbLinear	87	57	22	22	0.580	0.538	0.688	0.604	0.798	0.609	0.559	0.613
396	M9	a3	rose	logistic_reg	117	27	25	19	0.723	0.618	0.818	0.812	0.824	0.672	0.669	0.718
397	M9	a3	rose	decision_tree_gini	130	14	40	4	0.713	0.493	0.828	0.903	0.765	0.647	0.723	0.661
398	M9	a3	rose	decision_tree_entropy	133	11	41	3	0.723	0.489	0.836	0.924	0.764	0.648	0.711	0.663
399	M9	a3	rose	random_forest	92	52	13	31	0.654	0.625	0.739	0.639	0.876	0.679	0.571	0.682
400	M9	a3	rose	xgbTree	82	62	12	32	0.606	0.606	0.689	0.569	0.872	0.694	0.514	0.648
401	M9	a3	rose	xgbLinear	80	64	21	23	0.548	0.528	0.653	0.556	0.792	0.571	0.556	0.591
402	M9	a6	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.673	0.888	0.878
403	M9	a6	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
404	M9	a6	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
405	M9	a6	none	random_forest	140	4	38	6	0.777	0.693	0.870	0.972	0.787	0.657	0.785	0.781
406	M9	a6	none	xgbTree	138	6	38	6	0.766	0.642	0.863	0.958	0.784	0.659	0.878	0.752
407	M9	a6	none	xgbLinear	137	7	37	7	0.766	0.644	0.862	0.951	0.787	0.683	0.878	0.753
408	M9	a6	smote	logistic_reg	129	15	25	19	0.787	0.698	0.866	0.896	0.838	0.714	0.709	0.782
409	M9	a6	smote	decision_tree_gini	104	40	16	28	0.702	0.639	0.788	0.722	0.867	0.684	0.605	0.714
410	M9	a6	smote	decision_tree_entropy	122	22	35	9	0.697	0.534	0.811	0.847	0.777	0.718	0.688	0.672
411	M9	a6	smote	random_forest	108	36	24	20	0.681	0.588	0.783	0.750	0.818	0.673	0.584	0.685
412	M9	a6	smote	xgbTree	95	49	24	20	0.612	0.544	0.722	0.660	0.798	0.639	0.608	0.633
413	M9	a6	smote	xgbLinear	98	46	24	20	0.628	0.553	0.737	0.681	0.803	0.640	0.624	0.645
414	M9	a6	rose	logistic_reg	115	29	25	19	0.713	0.609	0.810	0.799	0.821	0.686	0.653	0.709
415	M9	a6	rose	decision_tree_gini	123	21	29	15	0.734	0.613	0.831	0.854	0.809	0.417	0.677	0.722
416	M9	a6	rose	decision_tree_entropy	102	42	28	16	0.628	0.530	0.745	0.708	0.785	0.587	0.640	0.637
417	M9	a6	rose	random_forest	85	59	13	31	0.617	0.606	0.702	0.590	0.867	0.683	0.545	0.654
418	M9	a6	rose	xgbTree	91	53	12	32	0.654	0.630	0.737	0.632	0.883	0.700	0.541	0.683
419	M9	a6	rose	xgbLinear	66	78	11	33	0.527	0.577	0.597	0.458	0.857	0.666	0.413	0.587
420	M9	a9	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.685	0.894	0.878
421	M9	a9	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
422	M9	a9	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
423	M9	a9	none	random_forest	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.643	0.842	0.878
424	M9	a9	none	xgbTree	139	5	40	4	0.761	0.610	0.861	0.965	0.777	0.586	0.874	0.736
425	M9	a9	none	xgbLinear	139	5	40	4	0.761	0.610	0.861	0.965	0.777	0.617	0.907	0.736
426	M9	a9	smote	logistic_reg	132	12	28	16	0.787	0.698	0.868	0.917	0.825	0.736	0.732	0.783
427	M9	a9	smote	decision_tree_gini	112	32	20	24	0.723	0.639	0.812	0.778	0.848	0.725	0.627	0.725
428	M9	a9	smote	decision_tree_entropy	101	43	26	18	0.633	0.545	0.745	0.701	0.795	0.574	0.625	0.645
429	M9	a9	smote	random_forest	99	45	15	29	0.681	0.630	0.767	0.688	0.868	0.717	0.579	0.699
430	M9	a9	smote	xgbTree	97	47	23	21	0.628	0.559	0.735	0.674	0.808	0.644	0.617	0.647

Table A.4: Model Performance Measures for Object Localization of Aim 4

	method	α	subsample.	model	TP	FP	FN	TN	acc.	b-acc.	F1	precision	recall	AUC	brier	prfm.
431	M9	a9	smote	xgbLinear	109	35	23	21	0.691	0.600	0.790	0.757	0.826	0.688	0.665	0.695
432	M9	a9	rose	logistic_reg	116	28	24	20	0.723	0.623	0.817	0.806	0.829	0.685	0.677	0.720
433	M9	a9	rose	decision_tree_gini	116	28	41	3	0.633	0.418	0.771	0.806	0.739	0.613	0.679	0.594
434	M9	a9	rose	decision_tree_entropy	136	8	43	1	0.729	0.435	0.842	0.944	0.760	0.650	0.699	0.639
435	M9	a9	rose	random_forest	91	53	16	28	0.633	0.598	0.725	0.632	0.850	0.669	0.575	0.662
436	M9	a9	rose	xgbTree	76	68	18	26	0.543	0.543	0.639	0.528	0.809	0.617	0.517	0.591
437	M9	a9	rose	xgbLinear	72	72	18	26	0.521	0.533	0.615	0.500	0.800	0.576	0.492	0.574
438	M9	all	none	logistic_reg	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.710	0.887	0.878
439	M9	all	none	decision_tree_gini	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
440	M9	all	none	decision_tree_entropy	144	0	43	1	0.771	0.885	0.870	1.000	0.770	0.489	0.827	0.878
441	M9	all	none	random_forest	144	0	44	0	0.766		0.867	1.000	0.766	0.716	0.832	
442	M9	all	none	xgbTree	143	1	39	5	0.787	0.810	0.877	0.993	0.786	0.712	0.891	0.843
443	M9	all	none	xgbLinear	141	3	35	9	0.798	0.776	0.881	0.979	0.801	0.734	0.896	0.828
444	M9	all	smote	logistic_reg	133	11	28	16	0.793	0.709	0.872	0.924	0.826	0.740	0.718	0.791
445	M9	all	smote	decision_tree_gini	115	29	33	11	0.670	0.526	0.788	0.799	0.777	0.429	0.696	0.657
446	M9	all	smote	decision_tree_entropy	117	27	35	9	0.670	0.510	0.791	0.812	0.770	0.596	0.671	0.650
447	M9	all	smote	random_forest	108	36	22	22	0.691	0.605	0.788	0.750	0.831	0.685	0.572	0.697
448	M9	all	smote	xgbTree	111	33	26	18	0.686	0.582	0.790	0.771	0.810	0.682	0.686	0.686
449	M9	all	smote	xgbLinear	112	32	30	14	0.670	0.547	0.783	0.778	0.789	0.622	0.697	0.665
450	M9	all	rose	logistic_reg	120	24	25	19	0.739	0.635	0.830	0.833	0.828	0.684	0.669	0.733
451	M9	all	rose	decision_tree_gini	114	30	23	21	0.718	0.622	0.811	0.792	0.832	0.391	0.693	0.717
452	M9	all	rose	decision_tree_entropy	98	46	16	28	0.670	0.619	0.760	0.681	0.860	0.650	0.605	0.689
453	M9	all	rose	random_forest	107	37	19	25	0.702	0.626	0.793	0.743	0.849	0.692	0.579	0.709
454	M9	all	rose	xgbTree	109	35	14	30	0.739	0.674	0.816	0.757	0.886	0.716	0.620	0.745
455	M9	all	rose	xgbLinear	122	22	27	17	0.739	0.627	0.833	0.847	0.819	0.706	0.713	0.730