RESEARCH-ARTICLE

Moving Beyond the Simulator: Interaction-Based Drunk Driving Detection in a Real Vehicle Using Driver Monitoring Cameras and Real-Time Vehicle Data

ROBIN DEUBER, Swiss Federal Institute of Technology, Zurich, Zurich, ZH, Switzerland

PATRICK LANGER, Swiss Federal Institute of Technology, Zurich, Zurich, ZH, Switzerland

MATHIAS KRAUS, University of Regensburg, Regensburg, Bayern, Germany

MATTHIAS PFÄFFLI, University of Bern, Bern, BE, Switzerland

MATTHIAS BANTLE, University of Bern, Bern, BE, Switzerland

FILIPE BARATA, Swiss Federal Institute of Technology, Zurich, Zurich, ZH, Switzerland

View all

Published: 26 April 2025

Citation in BibTeX format

CHI 2025: CHI Conference on Human Factors in Computing Systems
April 26 - May 1, 2025
Yokohama, Japan

Conference Sponsors:
SIGCHI

Open Access Support provided by:

Swiss Federal Institute of Technology, Zurich

University of St. Gallen

University of Bern

University of Regensburg

CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (April 2025)
https://doi.org/10.1145/3706598.3714007
ISBN: 9798400713941

Moving Beyond the Simulator: Interaction-Based Drunk Driving Detection in a Real Vehicle Using Driver Monitoring Cameras and Real-Time Vehicle Data

Robin Deuber
Department of Management, Technology, and Economics
ETH Zürich
Zürich, Switzerland
rdeuber@ethz.ch

Patrick Langer
Department of Management, Technology, and Economics
ETH Zürich
Zürich, Switzerland
planger@ethz.ch

Mathias Kraus
Faculty of Informatics and Data Science
University of Regensburg
Regensburg, Germany
mathias.kraus@ur.de

Matthias Pfäffli
Institute of Forensic Medicine
University of Bern
Bern, Switzerland
matthias.pfaeffli@irm.unibe.ch

Matthias Bantle
Department of Chemistry, Biochemistry and Pharmaceutical Sciences
University of Bern
Bern, Switzerland
matthias.bantle@irm.unibe.ch

Filipe Barata
Department of Management, Technology, and Economics
ETH Zürich
Zürich, Switzerland
fbarata@ethz.ch

Florian von Wangenheim
Department of Management, Technology, and Economics
ETH Zürich
Zürich, Switzerland
fwangenheim@ethz.ch

Elgar Fleisch
Department of Management, Technology, and Economics
ETH Zürich
Zürich, Switzerland
Institute of Technology Management
University of St. Gallen
St. Gallen, Switzerland
efleisch@ethz.ch

Wolfgang Weinmann*
Institute of Forensic Medicine
University of Bern
Bern, Switzerland
wolfgang.weinmann@irm.unibe.ch

Felix Wortmann*
Institute of Technology Management
University of St. Gallen
St. Gallen, Switzerland
felix.wortmann@unisg.ch

Abstract

Alcohol consumption poses a significant public health challenge, presenting serious risks to individual health and contributing to over 700 daily road fatalities worldwide. Digital interventions can play a crucial role in reducing these risks. However, reliable drunk driving detection systems are vital to effectively deliver these interventions. To develop and evaluate such a system, we conducted an interventional study on a test track to collect real vehicle data from 54 participants. Our system reliably identifies non-sober driving with an area under the receiver operating characteristic curve (AUROC) of 0.84 ± 0.11 and driving above the WHO-recommended blood alcohol concentration limit of 0.05 g/dL with an AUROC of 0.80 ± 0.10. Our models rely on well-known physiological drunk driving patterns. To the best of our knowledge, we are the first to (1) rigorously evaluate the potential of (2) driver monitoring cameras and real-time vehicle data for detecting drunk driving in a (3) real vehicle.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**; **Ubiquitous and mobile computing**; • **Applied computing → Consumer health**.

Keywords

health, safety, driving, eye movement, vehicle interaction, driver monitoring camera

*Joint last author.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3714007

ACM Reference Format:
Robin Deuber, Patrick Langer, Mathias Kraus, Matthias Pfäffli, Matthias Bantle, Filipe Barata, Florian von Wangenheim, Elgar Fleisch, Wolfgang Weinmann, and Felix Wortmann. 2025. Moving Beyond the Simulator: Interaction-Based Drunk Driving Detection in a Real Vehicle Using Driver Monitoring Cameras and Real-Time Vehicle Data. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3706598.3714007

1 Introduction

Alcohol represents a significant public health challenge. Approximately 5% of global deaths can be attributed to alcohol consumption. Research has shown that alcohol has an impact on 27 different causes of diseases, injuries, and fatalities (e.g., alcoholic cardiomyopathy, alcohol use disorder, cirrhosis of the liver) [122]. Consequently, the Sustainable Development Goals (SDG) address harmful alcohol consumption [110]. More specifically, target 3.5 calls for the mitigation of harmful alcohol use. Digital interventions (e.g., mobile apps to address irresponsible drinking) offer a viable pathway to address this objective [5, 48, 98]. However, to implement such interventions effectively, accurate prediction of the blood alcohol concentration (BAC) is essential.

With respect to mobility, alcohol-related road crashes result in more than 700 fatalities daily, accounting for 22% of fatal driving-related incidents globally [112]. In line with SDG target 3.5, target 3.6 aims to reduce mortality from road traffic accidents [110]. A potential solution could be driverless mobility. While predictions vary, experts concur that fully autonomous driving without any driver involvement is unlikely to be widely implemented in the coming decades [20, 38, 71]. Until complete autonomy is achieved, drunk driving remains 'the cancer of mobility' and a core issue to be addressed [112].

There is a growing political commitment to addressing the high incidence of drunk driving accidents. In the United States (US), for instance, Congress has mandated that future vehicles be equipped with in-vehicle drunk driving detection and prevention systems [111]. To implement this mandate, the US National Highway Traffic Safety Administration (NHTSA) issued an Advance Notice of Proposed Rulemaking (ANPRM) in December 2023 to collect input and feedback from all relevant stakeholders including the public. Thereby, NHTSA articulated their intention to enforce a system that operates "passively" (page 28 of [78]).

Unfortunately, today, reliable methods for determining driver alcohol intoxication involve in-vehicle ignition interlock breathalyzers. These devices are not considered "passive" under legislative definitions [78]. Furthermore, they also come with significant costs. Installation costs range from 70 to 200 USD, and monthly maintenance and calibration fees are between 60 and 100 USD [51, 88].

Therefore, a scalable and cost-effective approach for alcohol detection is still to be developed. Ideally, such a solution would utilize sensors already available in modern vehicles. These vehicles, for example, have a controller area network (CAN) bus system that facilitates communication among various vehicle components including sensors for controls (e.g., steering wheel angle sensor) and sensors monitoring vehicle state (e.g., vehicle speed). Furthermore, latest vehicles are equipped with driver monitoring camera (DMC) devices, which are mandated by laws in regions like the European Union (EU) [29] and are essential for achieving high ratings in safety tests across the world [2, 33, 74]. In this work, we develop and evaluate a machine learning (ML) system designed to detect drunk driving by utilizing DMC and real-time CAN data, which are standard features in modern vehicles.

1.1 Contributions

To develop and evaluate the drunk driving detection system, we conducted a randomized controlled trial involving $n = 54$ participants who drove a real car on a test track under varying BAC levels. Using this data, we derived a drunk driving detection model that performs two classification tasks: determining whether the driver is under the influence of alcohol, and whether the BAC exceeds the World Health Organization (WHO)-recommended limit of 0.05 g/dL. This work builds upon latest simulator findings [54] and aims to investigate, if these findings can be replicated leveraging a real car. More specifically, the goal is to analyze if existing simulator-based features and ML pipelines can be transferred into a real vehicle environment, and what adaptations might be necessary. In addition, the work is geared towards understanding the magnitude of potential performance degradations and their practical implications.

Contributions: (1) We demonstrate that the performance of our real vehicle drunk driving detection system is comparable to previous simulator-based findings, which rely on selected features not available in cars today. Additionally, our system does not leverage high quality lab cameras. Instead, we built upon lower quality video data that is captured by industry-standard driver monitoring cameras. (2) To cope with feature and camera limitations, we extended existing work and derived two additional types of camera features. Moreover, to match the performance of the simulator findings, we also included CAN data as a second modality. (3) We demonstrate strong generalization capabilities to unseen drivers who were not part of the ML training set. (4) The analysis of our predictive model shows that our classification relies upon well-known physiological patterns that are associated with drunk driving. (5) By incorporating two control groups, we address potential placebo effects and compensatory behaviors as well as influences such as training and drowsiness effects.

Novelty: To the best of our knowledge, we are the first to implement and rigorously evaluate a drunk driving detection system in real vehicles, utilizing DMC and real-time CAN bus data. Previous research has primarily focused on statistical analyses of drunk driving (e.g., [46, 53, 106]), detection systems have previously been developed based on simulator studies only (e.g., [52, 54, 59]).

Significance: As society and policymakers seek effective strategies to mitigate harmful alcohol consumption and prevent drunk driving, we offer a solution that leverages existing vehicle hardware to accurately predict a driver's BAC, serving as a foundation for successful digital interventions (e.g., mobile app interventions or vehicle-triggered safety measures). Committed to incremental science and replication, we have published the source code to foster further research and development in this field: https://github.com/im-ethz/CHI25_Drunk-Driving-Detection-in-a-Real-Vehicle.git

2 Related work

This section surveys critical research on alcohol detection and drunk driving interventions within the Human-Computer Interaction (HCI) field. The subsequent sections delve into the evolution of driver-state detection technologies and specific drunk driving detection methods.

2.1 Alcohol detection and harmful use prevention

Alcohol consumption and drunk driving are recognized as important public health concerns. The HCI community has explored a variety of approaches to address harmful drinking behaviors, both inside and outside vehicular contexts. In previous studies, participants' BAC was detected using data from smartphone activities during daily living, achieving an area under the receiver operating characteristic curve (AUROC) of 0.80 [58]. Similar approaches employing smartphone interactions have also been utilized in other research [68, 70]. SoberMotion combined smartphone technology with a breathalyzer to prevent individuals convicted of driving under the influence from operating vehicles while impaired [125]. Additionally, wearable sensors on wristbands have been employed for alcohol detection purposes [126]. The SoberComm program assists alcohol-dependent patients by facilitating the communication of alcohol-use data to family members and treatment teams [124]. Furthermore, other initiatives have aimed to support participants in reducing their risk to engage in harmful drinking through digital interventions [40, 65]. This prior research demonstrates that detecting alcohol use through human-computer interactions is feasible, and can serve as a basis for digital interventions that ultimately reduce harm.

2.2 Driver state detection

Today, driver state detection is core to active vehicle safety. While passive vehicle safety systems such as crumple zones aim to mitigate the effects of collisions and safeguard passengers, active vehicle safety systems such as emergency braking systems proactively avert accidents. Active safety systems fall into the category of advanced driver-assistance systems (ADAS), which encompass safety and comfort features. ADAS cover a wide range of interventions, from audiovisual warnings to full vehicle control. Thereby, driver state information is used to deploy, improve or personalize interventions [12, 31]. Consequently, detecting the driver's state becomes essential.

The predominant systems for driver state detection within modern vehicles focus on detecting driver drowsiness and distraction. They are typically categorized into three main types: physiological, DMC-based, and CAN data-based [49, 85]. Systems based on physiological parameters monitor biometric signals such as heart rate and brain activity to evaluate the driver's condition. DMC-based systems analyze the physical actions of the driver, including gaze direction, head movements, and facial expressions, using driver-monitoring cameras and computer vision technologies. Systems based on CAN data track changes in driving behavior, such as steering actions and lane-keeping.

Currently, both DMC and CAN-based systems individually account for approximately 44% each of the market's available drowsiness detection products, hence making up 88% of the total market [22]. However, DMCs become a standard feature in new vehicles. Their increasing availability and superior detection performance have led to rapid adoption in modern commercial drowsiness detection systems [22]. Furthermore, DMCs play a vital role in detecting distractions, such as texting [49].

Detection systems utilizing DMC-data typically fall into two categories. The first approach involves deriving human-interpretable features, such as the percentage of on-road versus off-road glances, to train a ML model [127]. The second, more recent approach, employs deep learning techniques, which use raw images directly as input for the ML model, as exemplified by Guo et al. [39].

2.3 Drunk driving detection

Extensive research has focused on the statistical analysis of drunk driving, examining both driver-vehicle interactions and physiological metrics of drivers [46, 47]. Studies conducted on test tracks also contribute to this body of knowledge. For instance, one study explored the effects of alcohol intoxication on eye movement metrics, utilizing diverse DMC configurations [53]. The findings indicated that higher levels of intoxication correlate with prolonged durations of glances, blinks, and fixations, and an increased focus of the driver's gaze on the road area. Another recent test track study demonstrated that alcohol impacts drivers' on/off-road gaze behavior [106].

Conversely, fewer studies have addressed the *detection* of drunk driving. To the best of our knowledge, no such study has been conducted on a test track. Existing studies have employed simulators (e.g., a seat and driver controls mounted on a fixed frame in front of a display [60], or a real but stationary car positioned in front of a wide-angle screen [62, 63, 102]), focusing on behavioral metrics such as pedal usage and steering, alongside simulated vehicle states like velocity. However, the rigor of these studies varies, and some face validity issues. For instance, some studies have employed goggles designed to simulate intoxication effects, rather than actual alcohol consumption [42, 43]. Moreover, the fidelity of the simulators varies. Some studies have incorporated driver-facing camera data, employing various camera system types (e.g., commercial eye-tracking devices [13, 14, 59], augmented with infrared and RGB cameras [52]). A major limitation of all the presented studies is their lack of testing on data of drivers not previously included in the training set of the classifier. The small participant pools and the use of the same drivers for both training and evaluation raise concerns about the generalizability of these methods to new data from unseen drivers.

To the best of our knowledge, only one detection method has been evaluated on out-of-sample subjects, proposed by Koch and Maritsch et al. [54]. It was developed from data gathered in a clinical simulator study with $n = 30$ participants [54]. It employed a sliding window approach for feature generation and logistic regression for making predictions. The proposed system successfully determined sobriety status with an AUROC of 0.88 and differentiated between participants below or above the WHO-recommended BAC limit of 0.05 g/dL with an AUROC of 0.79.

2.4 Importance of replication and real vehicle studies

The transferability of simulator findings to real-world vehicles remains a debated issue within the domain. While some studies affirm their validity (e.g., [55, 61]), others report inconsistent outcomes [47]. Such discrepancies are often attributed to the limited availability of comparable hardware and environmental variations in actual vehicles, potentially limiting simulator results' generalizability [129]. For example, differences in outcomes between simulators and actual vehicles are shown for semi-autonomous driving scenarios [83]. In sum, three principal factors are posited to underlie these validation challenges: (1) the varying degrees of realism in simulated environments, dependent on simulator fidelity [11, 19, 37]; (2) the absence of external influences encountered in real-world driving [61]; and (3) modified driver behavior due to the absence of genuine risk [27].

Previous literature has highlighted a replication crisis within the HCI field [25, 113], driven by a strong emphasis on novelty [119]. In response, the RepliCHI series has been organized annually from 2011 to 2014 [117–120] to foster incremental scientific progress. Our study builds upon the core ideas of RepliCHI, serving as a replication conducted on a test track that builds upon methodologies previously established. We conducted a conceptual replication study [120], closely adhering to the methodology employed in the study by Koch and Maritsch et al. [54], which we regard as the state-of-the-art and refer to as the "original study" herein. To enhance external validity, we implemented our study in a different environment (test track). Due to variations in sensor data availability in actual vehicles – for instance, industry-standard DMCs in place of high precision lab eye-tracking devices – we modified certain ML features to adapt to the new environment, while maintaining the core elements of the original study. This adjustment facilitates a more realistic application.

To our knowledge, this drunk driving detection study is the first conducted on a real test track with intoxicated drivers aimed at developing a drunk driving detection system using real-time vehicle and driver monitoring camera data. The study was conducted as an interventional clinical trial.

3 Data collection

We conducted a randomized, controlled, interventional single-center study (ClinicalTrials.gov NCT05796609), adhering strictly to the principles of the Helsinki Declaration, good clinical practice guidelines, Swiss health laws, and the Swiss ordinance on clinical research. This study was registered and received approval from the local ethics committee in Bern, Switzerland (ID 2022-02245). It took place between April 2023 and July 2023. All participants provided informed consent after receiving a comprehensive explanation of the procedures involved in the study. The following section outlines the methodology of our study.

3.1 Study design

We informed participants about the study's purpose, which included the potential administration of alcohol; however, we did not disclose the precise amount of alcohol administered or their target BAC. Alongside the treatment group, we involved a fully informed reference group that did not consume alcohol (open-label) and a placebo group that was not aware they received no alcohol (blinded). The treatment group ingested an alcoholic beverage, whereas the placebo group received a non-alcoholic placebo drink. When two participants attended the study simultaneously, our protocol stringently forbade them from discussing their estimated BAC.

Table 1: Study design.

Assignment	Masking	Treatment
Treatment group	single-blinded	alcoholic drink
Placebo group	single-blinded	placebo drink
Reference group	unblinded	-

Prior research on drunk driving supports the inclusion of placebo and reference groups in alcohol studies [36, 91, 105]. Despite this, the existence and significance of the placebo effect in such studies continue to be subjects of debate [34, 76]. In medical research, a placebo often demonstrates the effects of expectancy, typically leading to symptom alleviation due to the belief in the efficacy of a treatment [99]. Participants in these contexts usually have limited prior experience with the substance. In contrast, participants in this study brought previous alcohol experience, enabling them to rely on past experiences instead of forming new expectations. This background might lead to an inadvertent improvement in cognitive performance under placebo-controlled conditions if participants believe they are intoxicated [105].

Studies have employed blinded placebo groups using both within-subject [59] and between-subjects designs [13]. Additionally, other research has utilized an open-label sober reference condition in both within-subject [54, 62] and between-subjects formats [63]. The open-label reference condition aligns more closely with real-world scenarios, as drivers typically know whether they are sober or intoxicated. Our data collection study encompassed both conditions. First, we introduced a placebo control group to address potential placebo effects and compensatory behaviors, thus ensuring consistency with prior research. Next, we incorporated a reference control group to adjust for influences such as training (e.g., familiarization with the vehicle and the test track) and drowsiness. This group additionally compensated for environmental variations on the test track, such as changes in sunlight exposure throughout the day, while eliminating any confounding placebo effects. We present the details of these conditions in Table 1.

Following [30] and drawing on the design from the original study [54] and a related study [61], we determined the sample sizes of $n = 25$ for the treatment group and $n = 10$ for each control group. To account for potential dropout and technical failures, we adjusted the target sizes for the treatment and control groups to $n = 32$ and $n = 12$, respectively.

3.2 Participants

Eligibility criteria required participants to possess a driving license valid in Switzerland, be aged 21 or older, and have driven actively within the past six months. Further, they were required to have consumed alcohol at least occasionally during the same period. Exclusion criteria included health conditions that contraindicated alcohol consumption, use of medications or drugs that interact with alcohol, and excessive alcohol consumption.

Excessive alcohol consumption was assessed using two measures: (1) phosphatidylethanol (PEth) and (2) Alcohol Use Disorder Identification Test (AUDIT). PEth, an alcohol biomarker measured in a capillary blood sample, indicates chronic excessive alcohol consumption with values exceeding 200 ng/mL (further details in Section 3.3.1) [66, 93]. AUDIT, a 10-question screening tool developed by the WHO, detects hazardous or harmful alcohol consumption [90]. We excluded participants who scored 15 or higher.

Additional exclusion criteria included pregnancy, breastfeeding, or plans to become pregnant during the study. Participants with a history of excessive alcohol use, recent drug abuse, or a positive drug test were also excluded. Moreover, those unable to comply with study procedures due to language barriers, psychological disorders, or other impairments, as well as individuals who had participated in a drug study within the past 30 days, were deemed ineligible.

Of the 72 participants initially recruited, a total of 55 met the eligibility criteria. We excluded one participant from the analysis due to errors in the recorded CAN data. Their characteristics are summarized in Table 2. A comprehensive study flow diagram is provided in Appendix A.

3.3 Study procedure

3.3.1 Participant screening. We recruited participants through public announcements and advertisements. Interested individuals registered online, and we randomly selected participants for a phone screening. During the screening, the study team explained the study's purpose, answered participants' questions, and evaluated them against the inclusion and exclusion criteria. Eligible participants then attended a screening visit, where the study team collected capillary blood and urine samples to exclude excessive alcohol consumption by blood PEth concentrations [100], and to exclude pregnancy and drug abuse by urine tests.

The second visit served as the designated study day. To maintain standardized conditions across all participants, subjects were required to arrive in a fasted state (i.e., no food or caloric beverages for 4 hours). Initially, participants underwent a familiarization drive to get used to the study vehicle, the test track, and the procedure for receiving driving instructions to navigate the course. The actual baseline drive in a sober state, which later served as the negative class for classification, was then conducted first. Subsequently, the alcohol administration procedure commenced for the treatment group. The control group participants followed the identical procedure without receiving alcohol (either knowingly or unknowingly). To control for potential training effects, such as changes in driving behavior due to increased familiarity with the test track, the reference control group was introduced (see Section 3.1). While the control and placebo participants remained sober, the treatment group engaged in driving tasks at three different BAC levels.

3.3.2 Alcohol administration and measurement. The study design involved three distinct driving sessions for participants over the course of the study day. Participants in the treatment group completed one drive while sober, one drive with a BAC exceeding the WHO-recommended legal limit of 0.05 g/dL [121], and one drive with a BAC below this threshold. The nomenclature for these phases—*no alcohol*, *severe*, and *moderate*—is consistent with the terminology used in the original study [54].

Following the initial sober driving session, participants in the treatment group underwent an alcohol administration process aimed at achieving a target BAC of 0.08 g/dL. This protocol ensured consistency with the conditions of the original study. We calculated the amount of alcohol administered individually for each participant, taking into account their sex, weight, age, and height, using the Widmark formula [10, 115, 116]. Treatment group participants consumed a mixed drink consisting of vodka and orange juice, and their BAC was regularly monitored via breath alcohol concentration (BrAC) measurements. The severe driving phase commenced after participants' BAC peaked and subsequently dropped below 0.075 g/dL. Following the second driving session, participants took a break to allow their BAC to decrease. The moderate driving phase began once their BAC dropped below 0.035 g/dL. The participants' BAC are depicted in Figure 1.

We measured the participants' BAC using a professional breath alcohol measurement device. Until the debriefing, they were not informed of their BAC levels. We applied a conversion factor of 0.2 between BrAC and BAC, following Swiss regulations (e.g., a BAC of 0.08 g/dL corresponds to a BrAC of 0.4 mg/L) [26]. We used a Dräger 6820 device, which is officially licensed for use by public safety agencies in Switzerland [103]. The first measurement occurred 20 minutes after alcohol consumption to avoid distortion caused by residual alcohol in the mouth [64].

The control group participants followed the same temporal procedure. The duration of the break and the waiting period before the driving phases were determined by the average time required for participants in the treatment group, rather than by the participants' BAC. Participants in the placebo group received plain orange juice. Until the debriefing, they were not informed that their beverage was a placebo containing no alcohol. No study beverage was administered to the reference group.

During the initial screening, participants were informed that they might receive alcohol. However, they were blinded to their study group assignment, the target BAC, and the nature of the beverages provided. Following existing alcohol administration studies [41, 50], orange juice was mixed with vodka. Thereby, a bitter orange juice variant was chosen to mitigate the taste of vodka. To ensure the dominance of the orange flavor, the vodka was stored at -18 degrees Celsius. Drinks were dispensed to both treatment and placebo group participants in neutral bottles featuring a narrow outlet to minimize air exposure. The volume of the drinks varied from 3.0 dL to 5.5 dL, tailored to the calculated alcohol volume required. Placebo participants received a corresponding volume of beverage, devoid of alcohol.

3.3.3 Driving. A licensed driving instructor sat in the front passenger seat during the drives. The driving instructor issued commands to the participants, who were tasked with operating the vehicle as they would in real traffic, such as employing turn signals. Blinded to both the study group assignment and the participants' BAC, the driving instructor used dual pedals in case of emergencies to prevent injuries or damage. The driving instructor was tasked with intervening if participants overlooked or failed to stop at intersections

Table 2: Participant characteristics reported as mean ± standard deviation. AUDIT: Alcohol Use Disorder Identification Test; PEth: phosphatidylethanol.

	Treatment group	Placebo group	Reference group
n =	31	12	11
Sex (self-reported)	16 female, 15 male	6 female, 6 male	6 female, 5 male
Age [years]	37.5 ± 14.7	37.0 ± 17.0	36.5 ± 14.6
Weight [kg]	77.5 ± 15.3	75.3 ± 19.3	72.7 ± 12.6
Height [cm]	174.8 ± 8.0	170.5 ± 9.1	173.9 ± 10.1
Driver experience [years]	18.5 ± 14.5	17.2 ± 17.4	17.4 ± 14.3
Yearly driving distance [km]	14565 ± 12478	9000 ± 4848	9773 ± 6528
Occupational drivers	8	0	1
AUDIT score	4.81 ± 2.44	4.92 ± 2.02	3.82 ± 1.99
PEth			
Participants below detection limit 5 ng/mL	3	2	4
Participants below quantification limit 10 ng/mL	2	3	0
Quantifiable participants (above quantification limit 10 ng/mL) [ng/mL]	48.0 ± 43.2	61.0 ± 77.5	72.2 ± 70.3

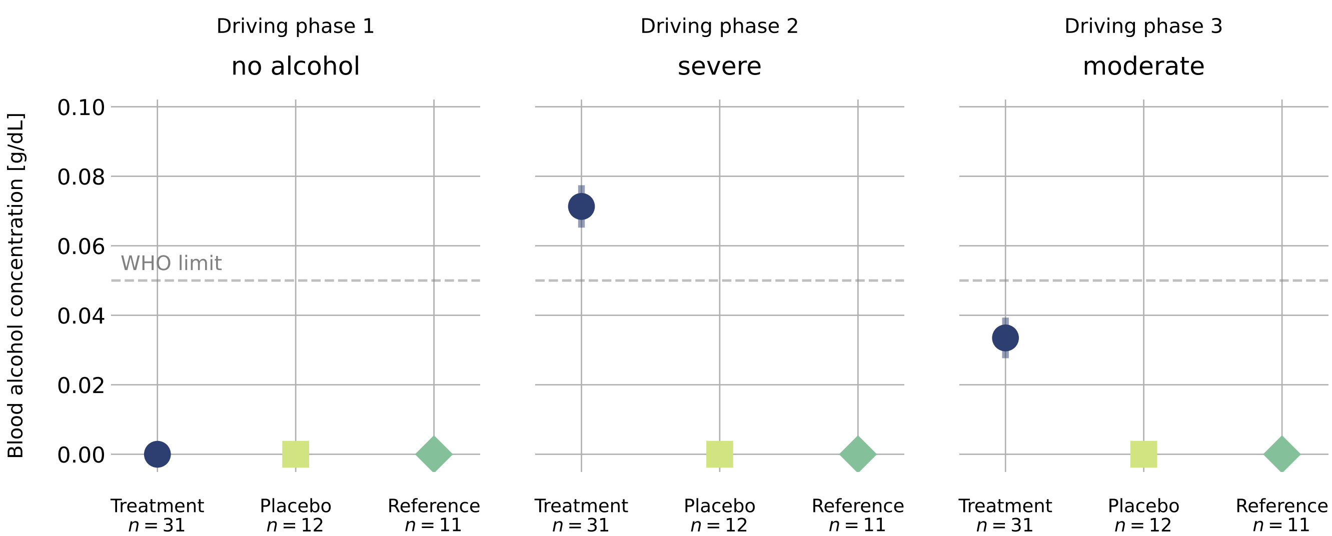

Figure 1: Each subplot corresponds to a driving phase, showing the average measured BAC for each study group. Error bars indicate standard deviation, and a dashed gray line marks the WHO-recommended BAC limit.

or crossroads. A study team member accompanied in the vehicle, supervising data collection and ensuring existing recordings.

Participants drove in three phases, each with different BAC levels (constant for the placebo and reference groups). In each phase, they drove three scenarios (highway, rural, and urban; further details in Section 3.5) in a random order, covering both travel directions, also randomly assigned. These directions were labeled with numbers (e.g., *urban 1* and *urban 2*). We briefly interrupted the drive between each scenario to measure participants' BAC.

3.4 Study vehicle and data recording system

The study vehicle was a 2020 Volkswagen Touran 1.5 TSI (Volkswagen AG, Wolfsburg, Germany) with an output of 110 kW. The vehicle featured an automatic transmission and was configured with 7 seats. Dual pedals were installed to allow vehicle control from both front positions.

We utilized two primary data sources: the data obtained from the DMC and the CAN bus communication recordings. The DMC employed was a near-infrared camera prototype (close to the commercially available version), featuring active illumination provided by Robert Bosch GmbH (Robert Bosch GmbH, Stuttgart, Germany).

This camera was mounted on the steering column behind the steering wheel, capturing images at a rate of 50 frames per second (Figure 2c). Comparable systems have been integrated by various car manufacturers into their vehicles, such as Volvo [97] and Toyota [108]. The camera's parameters were calibrated daily prior to the commencement of each study session.

Our study vehicle was also equipped with an embedded computer that interfaced with the vehicle's CAN bus. The data from both the CAN bus and the DMC were processed and stored on an in-vehicle computer.

3.5 Test track and scenarios

The study took place on a closed-off test track in Switzerland, which featured a variety of road structure types. The outer section of the track included broad curves and straight segments, while the inner section consisted of branched roads with sharp turns (Figure 2b [104]). The roads varied in width from 6 to 10 meters and were aligned with the surrounding terrain. In certain sections, middle lane markings were present, as shown in Figure 2a.

To replicate different driving conditions, we introduced artificial obstacles. We placed two stop signs at an intersection and added a crosswalk where participants were required to perform a full stop. Additionally, we simulated parked vehicles by placing cones on a section of the road.

We relied upon three distinct driving scenarios: highway, rural, and urban. In the highway scenario, participants drove at a maximum speed of 80 km/h, navigating minimal turns, large curve radii, and predominantly straight road segments, with no stops or obstacles encountered. The urban scenario allowed participants to drive up to 50 km/h and involved numerous turns with varying radii, as well as all the artificially introduced obstacles. The rural scenario combined elements from the other two, with participants driving at a maximum speed of 60 km/h and encountering a stop sign at an intersection and one obstacle that required them to maneuver around it. We developed these scenarios in collaboration with the Automobile Club of Switzerland [3].

4 Machine learning detection approach

To establish a model that is comparable to the original study [54], we initially focused on developing a classifier trained and evaluated exclusively on the data from the treatment group. In the subsequent phase, we integrated data from participants in the placebo and reference groups into both the modeling and evaluation processes.

In the following section, we provide a detailed description of our approach, which involves a sliding window combined with logistic regression [82]. The sliding window approach is frequently used for driver state detection [54, 61] but also for other tasks involving human interaction [44]. It offers several advantages: The sliding windows enable the capture of temporal dependencies within the time series data, while the use of human-interpretable aggregation functions ensures that interpretability remains high. This interpretability is further reinforced by the parsimonious logistic regression model, which maintains simplicity and clarity through coefficient analysis. An overview of our approach is depicted in Figure 3.

4.1 Feature generation

4.1.1 Input data. For our detection approach, we utilized data from two sensor modalities: DMC and CAN bus data. Unlike the original simulator study [54], where CAN signals contributed minimally to performance and were hence not included in the core analysis, we include both sensor modalities by default. Another reason to include both modalities is that we employ an industry-standard DMC in the field, rather than high precision lab eye-tracking devices. Therefore, our analysis incorporates data from both modalities, and we conduct an ablation study to evaluate the performance of each modality independently.

We employed a proprietary industry-standard algorithm, provided by Robert Bosch GmbH (Robert Bosch GmbH, Stuttgart, Germany), to extract head and gaze features from the driver's interactions. This algorithm performed several key functions: it detected the driver's face, localized the eye region, detected the pupil, and computed the gaze vector (Figure 2d). For head features, it identified facial landmarks and calculated the head pose. The outputs included the positions of the eyes, gaze orientation, and a binary eye state (open or closed). Additionally, the algorithm provided comprehensive head pose data, encompassing both position and orientation. The recording frequency of the DMC may fluctuate during load peaks on the study computer. Thus, to ensure data consistency across the analysis, we resampled this data to a fixed frame rate of 50 Hz. If required, missing values underwent interpolation during resampling. We applied linear interpolation for numerical data and nearest-neighbor interpolation for categorical data. This process was restricted to instances involving up to five consecutive missing entries, not exceeding 100 milliseconds. Any gaps exceeding five consecutive missing values were omitted.

We extracted controls signals (brake pedal pressure, gas pedal position, steering wheel angle, and steering wheel rotational velocity) and vehicle dynamics signals (longitudinal velocity, yaw velocity, and longitudinal and lateral acceleration) from the CAN bus data. In the same way as for the DMC data, we resampled these signals to a fixed frame rate of 50 Hz to maintain consistency across all collected data.

4.1.2 Data pre-processing. The head and eye data, resampled to 50 Hz, formed the foundation for computing various groups of derived signals. A comprehensive list of all features derived from this data is available in the Appendix B.

The first feature group comprises periodically sampled (1) *head movement* features, which include both linear and rotational velocities and accelerations. Each type of movement is represented by three spatial components along with a combined magnitude, providing a detailed characterization of head dynamics. We chose to exclude head position and orientation from our analysis due to their high individual variability.

The second group of features encompasses periodically sampled (2) *eye state* indicators, represented by a single binary signal for each eye. These signals specify whether the left and right eyes are open or closed, providing information on the blink patterns.

The third group of features consists of periodically sampled (3) *eye movement* characteristics. We specifically focused on gaze orientation, which is characterized by azimuth and elevation angles. From these, we calculated the velocity and acceleration in both

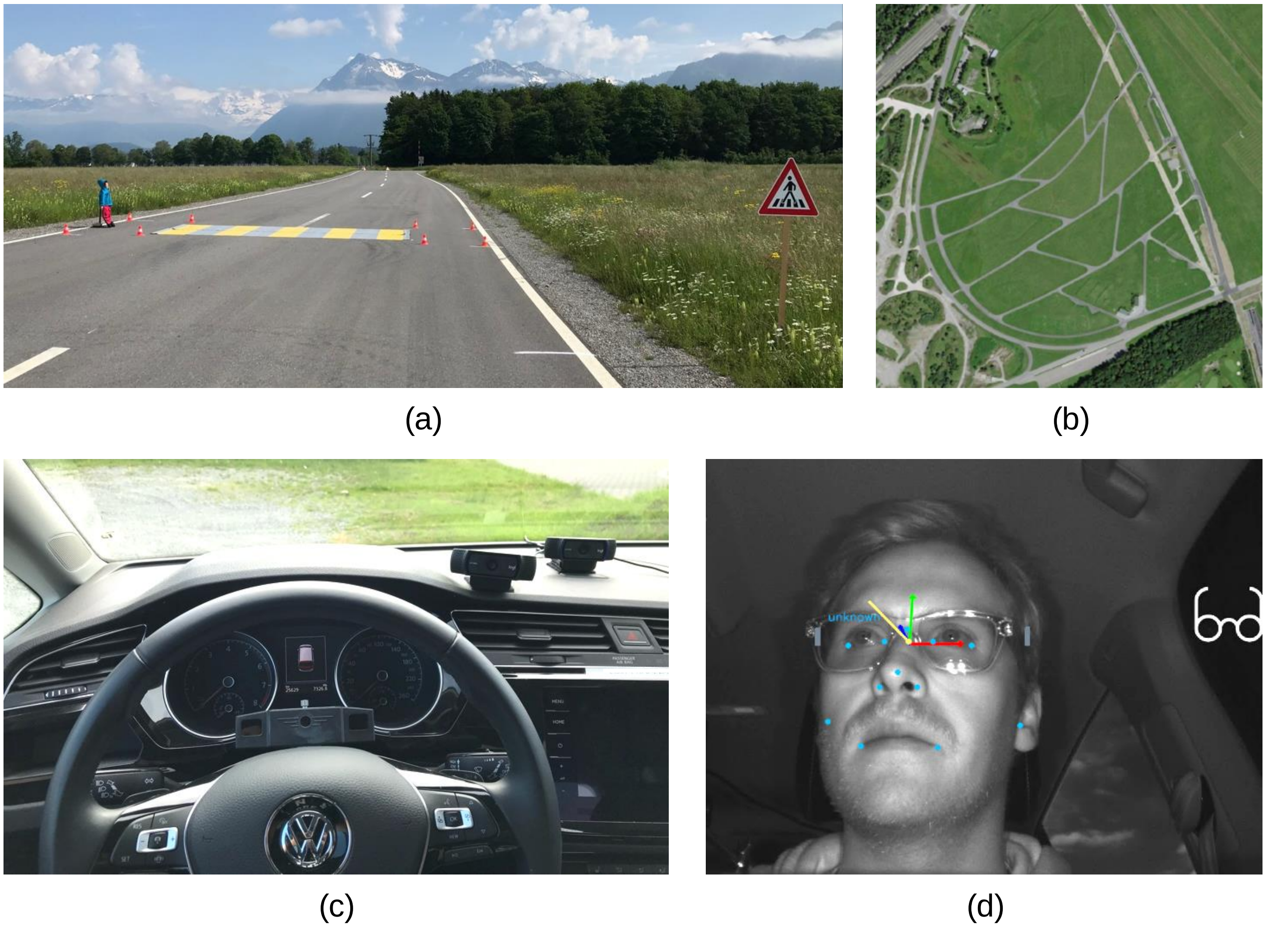

Figure 2: The figures display the following: (a) the test track featuring a crosswalk where participants were required to stop; (b) a top-down perspective on the test track (©swisstopo); (c) the driver monitoring camera (DMC) mounted on the steering column; and (d) the output from the DMC, displaying detected facial landmarks (blue dots), the mid-eye coordinate system, and the detected gaze direction (orange arrow).

directions, as well as their combined magnitude. Additionally, we introduced the gaze movement angle to capture the spatial relationship between changes in azimuth and elevation. We defined the gaze movement angle as measure between a horizontal reference line and a line connecting the initial and final gaze positions.

The fourth group of features involves (4) *gaze events*, classified using the REMoDNaV algorithm (robust eye-movement classification for dynamic stimulation) [17]. We utilized classified fixations and saccades. Fixations represent periods where the gaze remains relatively stable at a specific point, and saccades are rapid movements between fixations. REMoDNaV identifies fixations and saccades using a velocity-based algorithm and applies duration thresholds to the event candidates. We retained the REMoDNaV default thresholds for the fixation and saccade classification: minimum fixation duration of 40 milliseconds and minimum saccade duration of 10 milliseconds. The REMoDNaV algorithm not only identifies the duration of each event but also provides detailed metrics for each, including event amplitude, peak velocity, mean velocity, and median velocity.

The fifth group of features includes (5) *region-specific gaze events*, which are crucial for understanding the impact of alcohol consumption on driver attention. Recognizing that different gaze regions hold varying relevance for driving tasks, we assigned classified fixations to one of ten predefined gaze regions, each calibrated specifically to the study vehicle. These regions encompass the windscreen on both the driver and passenger sides, the left and right exterior mirrors, the interior mirror, the left and right front windows, the driver instruments, the navigation display, and the middle instrument cluster. This categorization facilitates a detailed analysis of how gaze behavior shifts across different areas of interest. Prior studies have demonstrated variations in gaze patterns across specific regions as a result of alcohol consumption [96, 106].

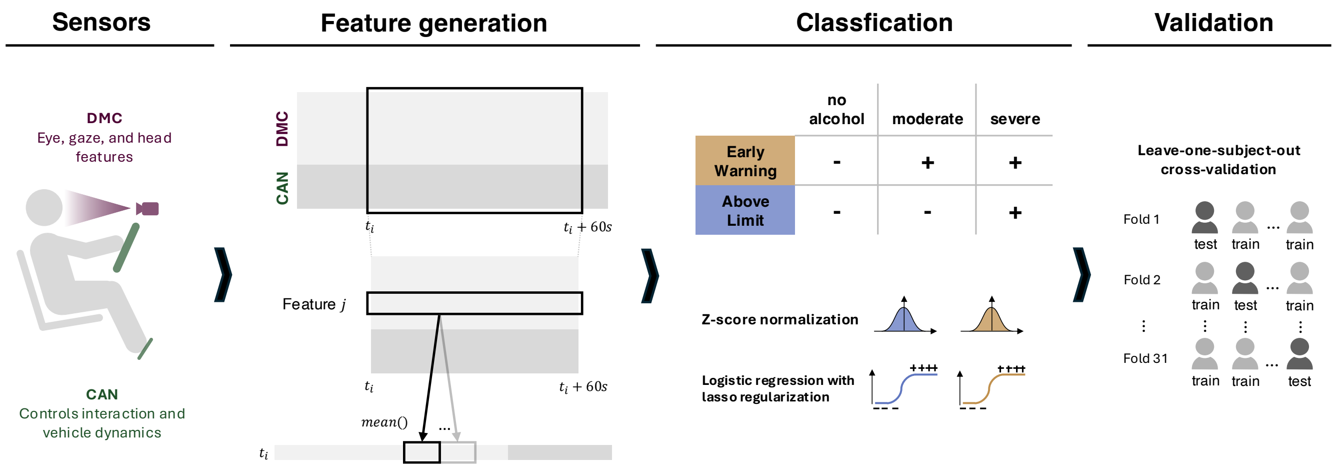

Figure 3: The figure displays an overview of our machine learning (ML) system designed to perform two classification tasks. DMC: driver monitoring camera; CAN: controller area network.

In relation to the CAN-based features, we distinguished between two primary types: (6) *controls interaction* and (7) *vehicle dynamics*. Beyond standard measures like position, velocity, and acceleration, we also calculated jerk — a critical signal in the automotive industry. Defined as the third derivative of position or the derivative of acceleration, jerk quantifies the rate of change of acceleration. It is widely used to evaluate ride comfort and vehicle stability [18, 45]. Consequently, for the CAN-based signals, we calculated velocity, acceleration, and jerk when these values were not pre-provided, ensuring comprehensive analysis capabilities for assessing vehicle behavior.

4.1.3 Sliding window and feature calculation. To extract features that encapsulate the temporal dynamics of time-series data, we employed a sliding window approach alongside a suite of aggregation functions. A window size of 60 seconds with a 1-second step size was chosen, as this configuration has proven effective in recent studies on impaired driving detection, such as the original simulator drunk driving detection [54], driver hypoglycemia detection [61], and drowsiness detection [128]. Consequently, we applied the aggregation functions over time-series segments comprising 3,000 samples (50 Hz * 60 seconds). Windows containing less than 75% of the expected samples were excluded (i.e., at the beginning or end of driving sequences). This approach strikes a balance between maximizing data inclusion and minimizing the incorporation of distorted data.

We combined various statistical aggregation functions to capture temporal dynamics, enhance robustness, and maintain interpretability. Specifically, mean and median were used to measure central tendency, while standard deviation, interquartile range, and 0.05 / 0.95 quantiles served as measures of dispersion. Minimum and maximum values were deliberately excluded due to their sensitivity to outliers. To further characterize the data distribution within each window, we calculated skewness and kurtosis. Additionally, power (mean square value) and the number of sign changes were computed, with the latter particularly motivated by the characteristics of CAN-bus-based controls interaction signals, as prior research in related domains of impaired driving suggests that the frequency of micro-corrections decreases in drowsy drivers [57]. In total, we generated 580 features.

4.2 Predictive modeling

In line with the original simulator study, we conducted two binary classification tasks: Early Warning and Above Limit. The objective of the Early Warning task was to determine whether participants had a BAC exceeding 0.00 g/dL (positive class). Accordingly, we labeled data from the treatment group participants' first phase as negative, while data from both the second and third phases were labeled as positive. The Above Limit task aimed to identify whether a participant's BAC surpassed the WHO-recommended limit of 0.05 g/dL (positive class). Thus, we labeled data from the treatment group participants' first and third phases as negative, and data corresponding to a BAC above the WHO limit as positive. Subsequently, we standardized the features using z-score normalization, transforming them to have a mean of 0.0 and a standard deviation of 1.0 based on the training data's distribution.

Our logistic regression model's strength lies in its intrinsic interpretability through coefficient analysis. We utilized Lasso (L1) regularization to enable the model to select relevant features while reducing others to zero, minimizing overfitting and enhancing feature interpretability. We set the (inverse) regularization coefficient to the default value of 1.0, utilized log loss (binary cross-entropy) for model optimization, and maintained the decision threshold at the standard 0.5. We applied balanced weights to accommodate class imbalances. The implementation was performed using Python 3.11.9 and Scikit-learn 1.4.2 [81].

4.3 Model evaluation

Evaluating the model's generalizability to unseen drivers is essential for its practical value. We employed a leave-one-subject-out (LOSO) cross-validation method [35], widely used in HCI research (e.g., [1, 15, 114]), iterating over all participants, training the model on $n - 1$ participants, and evaluating it on the remaining participant. This process was repeated for each participant, and performance metrics were reported as a macro-average, presented as *mean ± standard deviation*.

Our primary performance evaluation metric was the AUROC [16], chosen for its threshold-agnostic nature and its ability to report across all possible classification thresholds. This metric effectively summarizes the model's discrimination ability with a single value and is well-suited to handle imbalanced datasets [4]. Additionally, we reported the area under the precision-recall curve (AUPRC), balanced accuracy, and F1 score (a balanced combination of precision and recall), consistent with the original simulator results [54].

To further validate our proposed system, we analyzed its robustness against various design choices and parameter adjustments. Specifically, we examined how performance varied with different window sizes, ranging from 5 to 300 seconds, in the sliding window approach. Additionally, we evaluated the individual performance of different feature groups to understand the impact of each. We also investigated how the choice of different classifiers influenced the overall performance.

4.4 Control groups integration

To enhance the validity of our drunk driving detection approach, we incorporated data from the placebo and reference group participants, applying the same preprocessing and feature generation steps as for the treatment group. The placebo and reference groups were labeled as sober.

We employed LOSO cross-validation again, training the model on $n = 53$ participants and testing it on the remaining participant, aiming to maintain comparable evaluation metrics. Since the sensitivity (true positive (TP) rate) is not defined for placebo and reference groups, we calculated the AUROC and AUPRC for the treatment group only. We replaced balanced accuracy with standard accuracy, enabling calculation across all participants, with the accuracy for control group participants reflecting specificity (true negative (TN) rate). The weighted F1 score, defined even for the control group participants, ensured the overall metric remained computable as weights for the undefined F1 score of the true-labeled class defaulted to zero.

5 Results

In the following, we present the results of our study organized into general performance evaluations, the impact of sensor modalities, and a sensitivity analysis to validate the robustness of our ML models. The section concludes with an analysis of the feature contributions and the results of incorporating the control groups.

5.1 Performance evaluation

The dataset consists of 148 596 samples for both classification tasks. For Early Warning, 66% of the samples belong to the positive class (i.e., labeled as non-sober), while for Above Limit, 33% of the samples are in the positive class (i.e., labeled as above the WHO-recommended limit of 0.05 g/dL). Figure 4a displays the performance of our drunk driving detection system, utilizing a combined DMC- and CAN-based approach. The Early Warning model, which assesses whether participants are sober, achieves an AUROC of 0.84 ± 0.11. Conversely, the Above Limit system, which determines if a driver's BAC exceeds the WHO-recommended limit, records a slightly lower AUROC of 0.80 ± 0.10, with a comparably low standard deviation across all test participants.

Additional performance metrics are detailed in Figure 5 ("DMC + CAN"). The AUPRC exhibits variability: Early Warning achieves 0.91 ± 0.07, while Above Limit scores 0.67 ± 0.14, with the lower performance possibly attributed to the fewer positive samples in Above Limit. Both tasks display relatively similar balanced accuracy and F1 score, with Early Warning recording a balanced accuracy of 0.74 ± 0.10 and an F1 score of 0.75 ± 0.12, while Above Limit achieves 0.69 ± 0.10 for both metrics.

Figure 6 presents the confusion matrices for the two classification tasks, illustrating the proportion of each predicted class matching the true classes. Early Warning achieved relatively low rates of false positive (FP) and false negative (FN). Specifically, the FN rate for moderately intoxicated individuals was higher at 30% compared to 17% for severely intoxicated drivers. This difference likely arises from the less pronounced symptoms in moderately intoxicated drivers compared to those who are severely intoxicated. For both moderate and severe intoxication levels, Above Limit exhibited marginally higher FP and FN rates. The proportion of FP among sober participants remained low at 16%.

Figure 7 illustrates the performance of the two models across various driving scenarios, highlighting their consistent stability. The Early Warning model excels in rural environments but shows decreased performance in urban settings. Conversely, Above Limit achieves its best performance in urban scenarios and its lowest on highways. Overall, Early Warning tends to outperform Above Limit slightly.

5.2 Comparison of DMC and CAN approaches

Figures 4b and 4c display the performance outcomes when employing DMC- and CAN-based features individually. Utilizing a single sensor modality generally results in lower performance. Combining the two modalities yields an AUROC of 0.84 ± 0.11 for Early Warning, whereas DMC achieves 0.79 ± 0.12 and CAN registers 0.75 ± 0.13. Similarly, for Above Limit, the combined modalities achieve an AUROC of 0.80 ± 0.10, with DMC at 0.75 ± 0.10 and CAN at 0.72 ± 0.10. This pattern persists across both classification tasks: the combination of modalities performs best, followed by DMC, with CAN exhibiting the lowest performance. For Early Warning, the standard deviation in performance across participants mirrors this trend, whereas for Above Limit, it remains consistent.

Figure 5 illustrates that the following two patterns are evident across all metrics: (1) the combined modality approach outperforms single modality methods, and (2) the performance metrics for Early Warning surpass those for Above Limit. Specifically, the AUPRC for the combined approach registers 0.91 ± 0.07 for Early Warning and 0.67 ± 0.14 for Above Limit. The DMC-only method achieves an AUPRC of 0.88 ± 0.08 for Early Warning and 0.59 ± 0.13 for

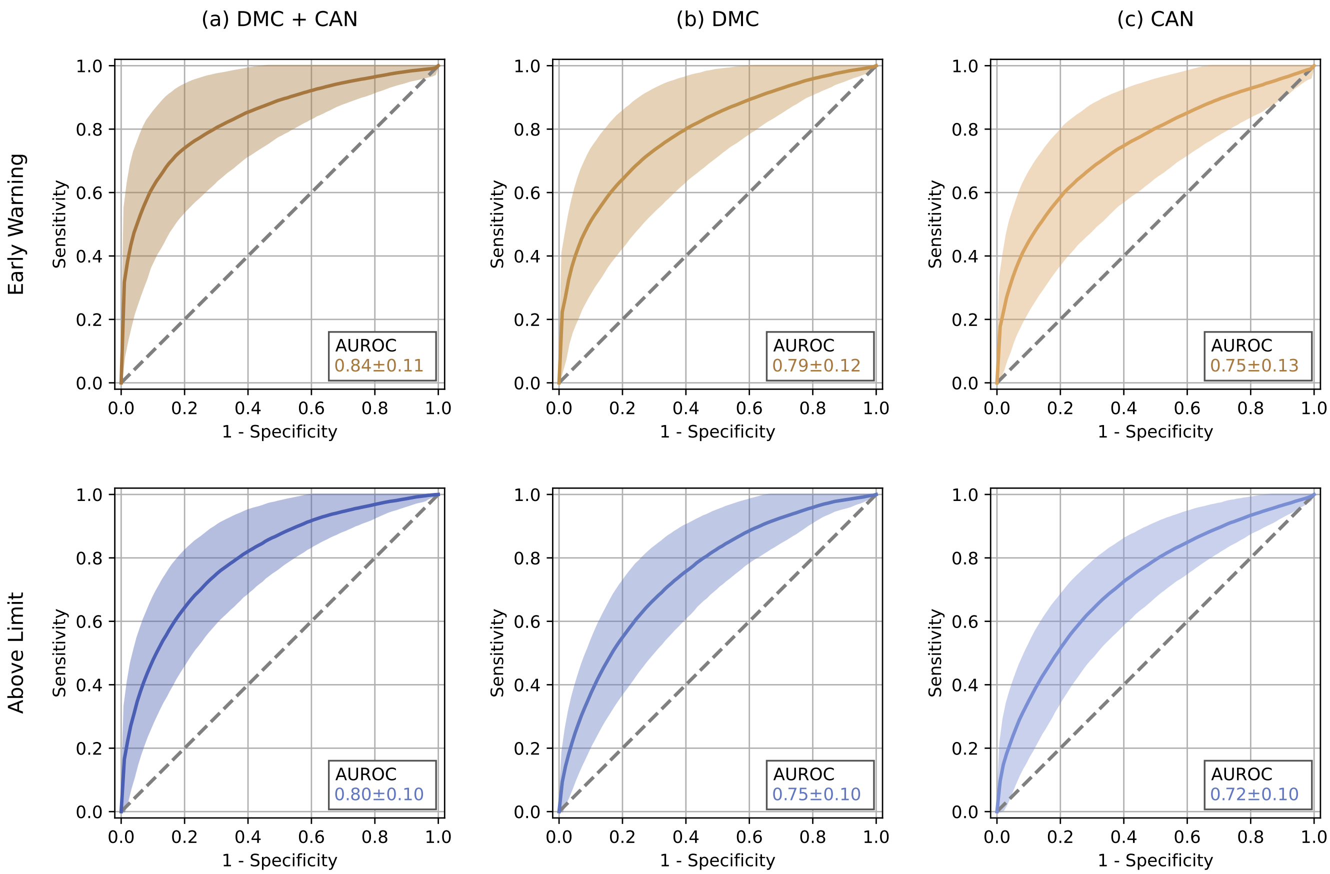

Figure 4: The figure displays subplots of receiver operating characteristic curves, averaged across all participants. The shaded areas represent the standard deviation among participants. The performance is reported as the area under the receiver operating characteristic curve (AUROC). A dashed gray line in each subplot represents the performance of a random classifier. Each row corresponds to a distinct classification task: Early Warning and Above Limit. Column (a) illustrates the combined performance of driver monitoring camera (DMC) and controller area network (CAN) sensors. Columns (b) and (c) show the individual performances of DMC and CAN sensors, respectively.

Above Limit, while CAN scores 0.85 ± 0.09 for Early Warning and 0.57 ± 0.13 for Above Limit. Our proposed approach yields a balanced accuracy of 0.74 ± 0.10 for Early Warning and 0.69 ± 0.10 for Above Limit. In a DMC-only setting, we record 0.69 ± 0.10 for Early Warning and 0.66 ± 0.08 for Above Limit. Utilizing only CAN, the performance measures at 0.66 ± 0.10 for Early Warning and 0.63 ± 0.08 for Above Limit. When both modalities are employed, the weighted F1 score reaches 0.75 ± 0.12 for Early Warning and 0.69 ± 0.10 for Above Limit. DMC attains 0.70 ± 0.12 for Early Warning and 0.66 ± 0.09 for Above Limit. The CAN-only strategy achieves 0.64 ± 0.12 for Early Warning and 0.62 ± 0.09 for Above Limit.

5.3 Robustness analysis

Figure 8 demonstrates the robustness of our detection system against various design and parameter changes. First, we evaluated different window sizes used in the sliding window approach (Figure 8a). As anticipated, the AUROC improved with increasing window sizes, as they captured more information; however, the rate of improvement diminished beyond a certain point. We chose a window size of 60 seconds beforehand based on prior research ([54, 61, 128]), balancing performance with the need to minimize the delay in detection. Second, we trained classifiers using each feature group individually (Figure 8b). In general, the feature groups provided complementary information, as their individual performances were lower compared to when combined. Third, we compared the results across different model choices (Figure 8c). Both linear models (logistic regression with lasso, ridge, or elastic net regularization) and non-linear models (support vector machine with radial basis function (RBF) kernel, random forest, gradient boosting, and multi-layer perceptron) exhibited relatively stable performance, with the random forest classifier yielding the lowest performance. The similarity in performance across models indicates

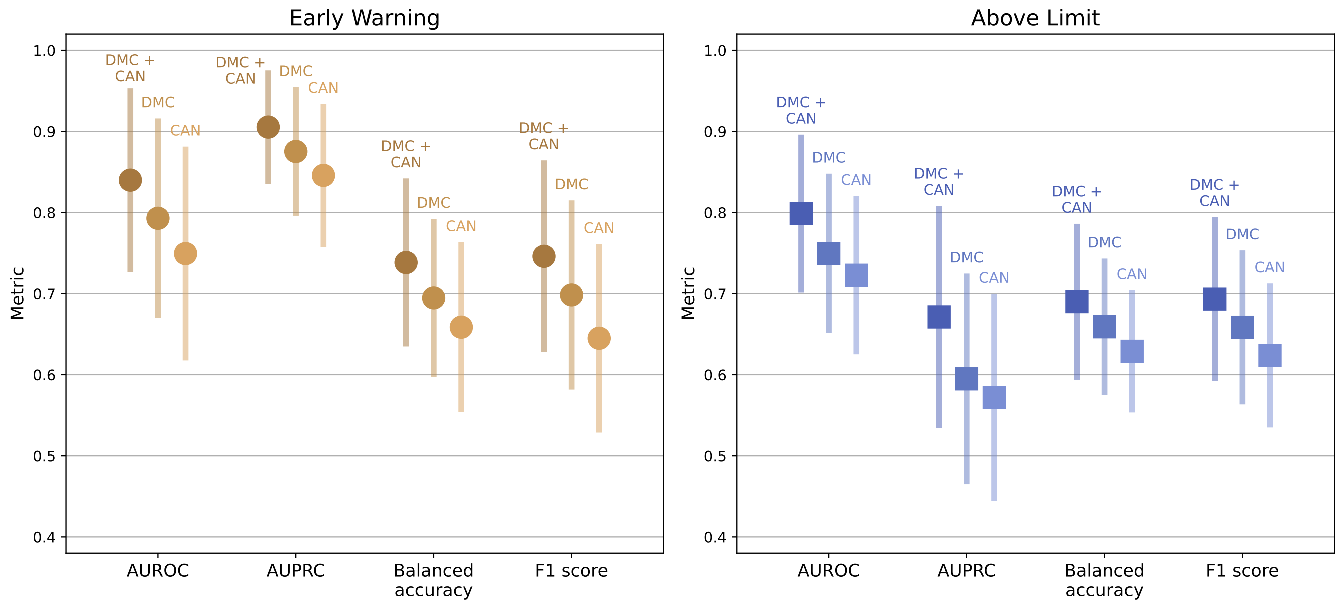

Figure 5: Each subplot corresponds to a classification task, illustrating the performance metrics area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), balanced accuracy, and F1 score (weighted by class). Each metric is evaluated for the three approaches: the combined driver monitoring camera (DMC) and controller area network (CAN) data model, and the individual DMC and CAN based models. Error bars represent the standard deviation across participants.

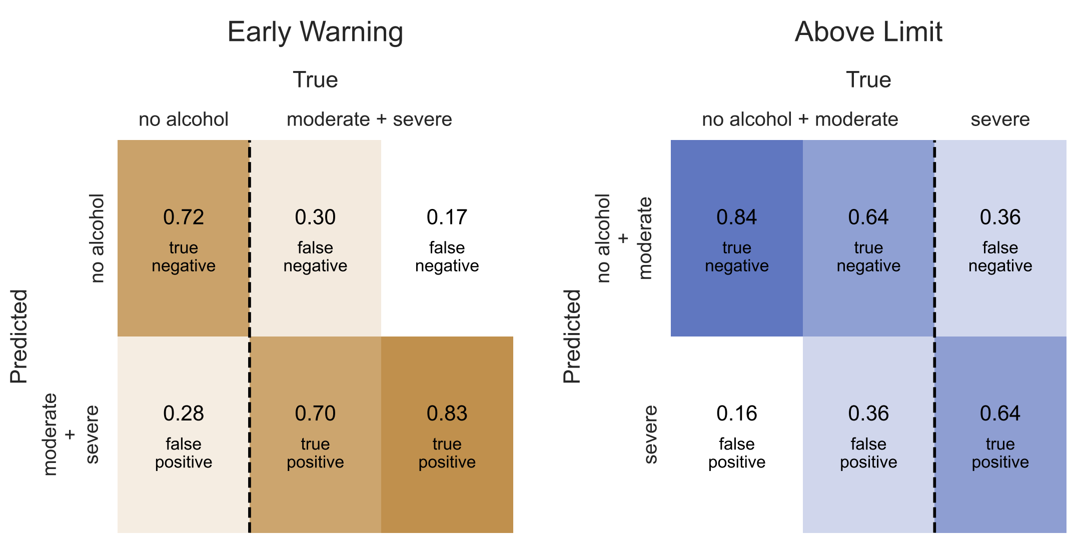

Figure 6: Each subplot displays the confusion matrix for one classification task's model. The confusion matrices show the relative frequencies of predicted classes (horizontal axis) compared to actual classes (vertical axis).

that our design decisions were robust. In general, in nearly all cases, Early Warning performed better than Above Limit.

5.4 Interpretability

To analyze the impact of each feature, or feature group respectively, we conducted several analyses. First, we calculated the sum of the magnitude of coefficients normalized by the total sum, as shown

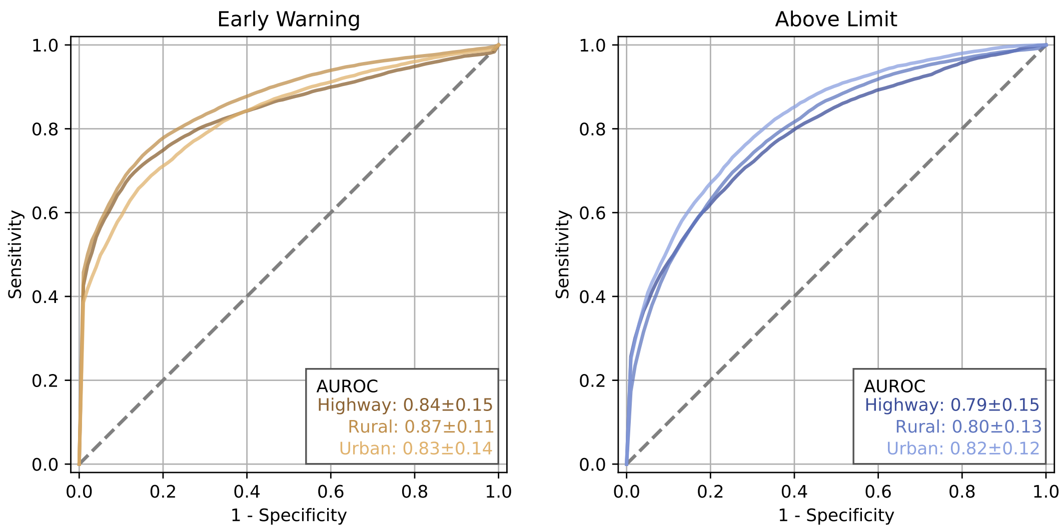

Figure 7: The figure displays subplots of receiver operating characteristic curves for different scenarios, averaged across all participants. The performance is reported as the area under the receiver operating characteristic curve (AUROC). A dashed gray line in each subplot represents the performance of a random classifier. Each subplot corresponds to a distinct classification task: Early Warning and Above Limit.

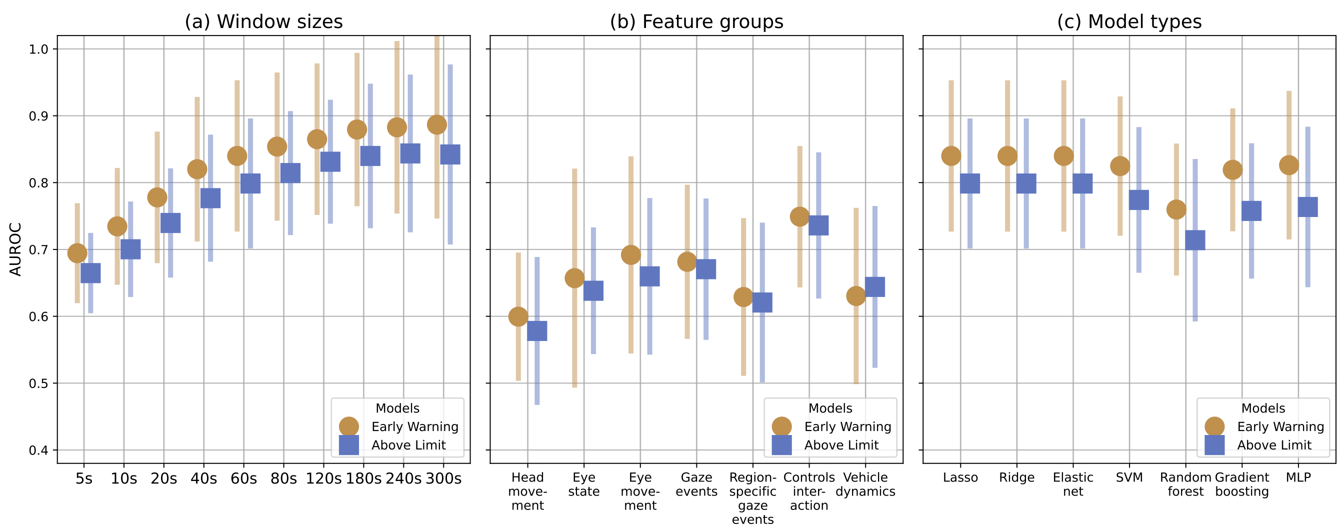

Figure 8: The figure presents our sensitivity analysis, where each subplot varies a specific parameter or design decision to examine its impact on performance. Performance is quantified using area under the receiver operating characteristic curve (AUROC). Error bars represent the standard deviations. The subplots detail different modifications: (a) varying window sizes, (b) using only one feature group, and (c) employing different classifiers.

in Figure 9. Head movement (31% Early Warning, 20% Above Limit), eye movement (18% Early Warning, 17% Above Limit), gaze events (18% Early Warning, 18% Above Limit), and controls interaction (21% Early Warning, 16% Above Limit) emerged as the feature groups with coefficient shares greater than 15%. Second, we evaluated the performance of models trained exclusively on a single feature group (Figure 8b). Notably, while head movement features had a high share of the absolute coefficient sum, the model relying solely on this feature group performed poorly. This discrepancy could potentially be linked to the large number of features within this group. Third, we examined the top three features within the groups of eye movement, gaze events, and controls interaction

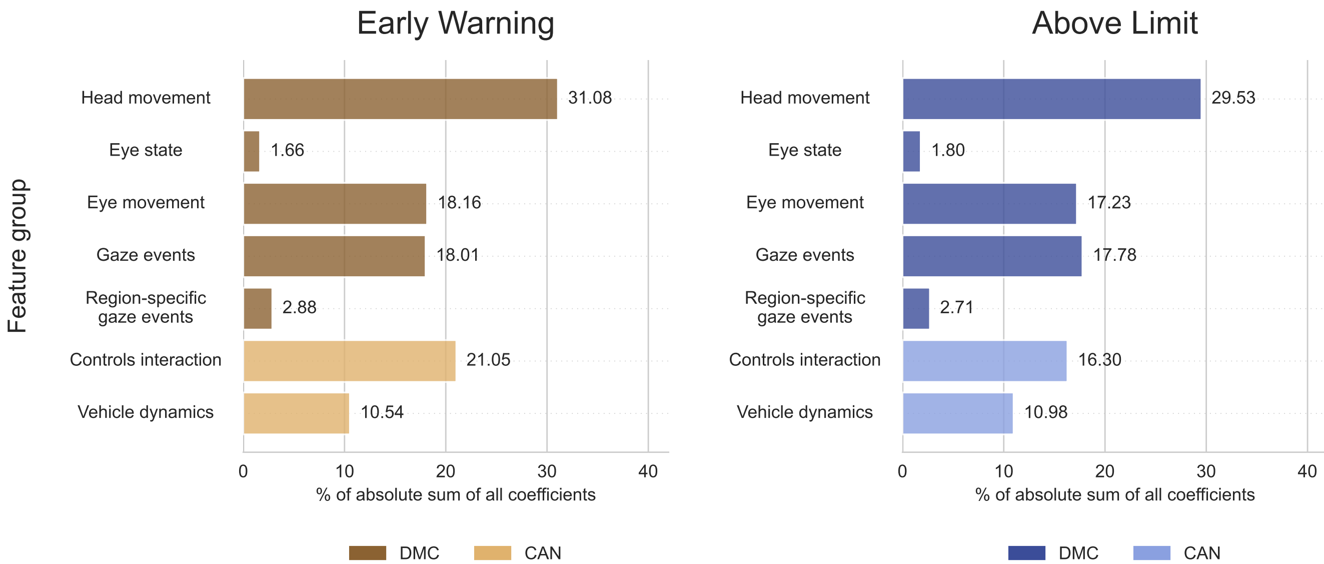

Figure 9: The figure illustrates the importance of each feature group. Each subplot corresponds to a classification task's model. The sum of the absolute values of the coefficients is calculated for each group and displayed as the percentage of the total sum.

(Figure 10), as these groups appeared most influential in the prior analyses. For instance, in the Early Warning model, the likelihood of predicting drunk behavior increased as the mean amplitude of saccades decreased, holding other factors constant (ceteris paribus).

5.5 Control group incorporation

In this section, we detail the outcomes achieved by incorporating both the treatment and the two control groups in our model's training and evaluation phases. We report the AUROC and AUPRC only for participants in the treatment group, as these metrics were not defined for control group participants. Figure 11 illustrates the performance of the two classifiers trained across all participant groups but assessed on those from the treatment group. For Early Warning, the AUROC recorded was 0.80 ± 0.11, being slightly lower compared to using solely treatment group participants for training, where it was 0.84 ± 0.11. For Above Limit, the average AUROC remained constant, though the standard deviation slightly reduced (0.80 ± 0.09, compared to 0.80 ± 0.10 with only treatment participants). Figure 12a displays additional metrics. As previously noted, the AUPRC shows great variation. The Early Warning accuracy and F1 score deteriorated marginally. Conversely, for Above Limit, both accuracy and F1 score not only outperformed those of Early Warning, but also surpassed the metrics when training involved only treatment group participants. Figure 12b depicts the AUROC across different aggregation window sizes, noting that performance improves with larger window sizes, though the diminishing returns effect is less pronounced than in models trained solely on the treatment group.

6 Discussion

In the following, we discuss the results of our study, exploring the implications and comparing them to existing literature. We discuss the interpretability of our model and assess the effectiveness of our system in realistic scenarios. We conclude with a discussion on the contributions of our study to public health and traffic safety, the limitations we encountered, and potential avenues for future research.

6.1 Post-hoc interpretation

Our method, which employs logistic regression, provides the advantage of allowing straightforward interpretation through coefficient analysis. In the following, we detail the most relevant coefficients and their interpretations. In general, research has indicated that drunk drivers exhibit diminished psychomotor skills, impaired perception, and divided attention [69, 73]. Our model captured these symptoms: For example, we observed altered saccadic eye movements under the influence of alcohol, characterized by a reduction in the number of saccades [32], slower movements [79, 86], and decreased amplitudes [86], coupled with increased variability [92]. Additionally, fixation durations increased, corroborating prior studies [72, 96]. These patterns were consistent with those found in the original study [54] and other previous test track studies [53]. Analyzing interactions with vehicle control elements revealed that braking, accelerating, and steering were all impacted. Literature characterizes the actions of drunk drivers as more aggressive [101] and highlights their deficits in maintaining proper longitudinal and lateral positions [84]. The features identified by our model, such as the standard deviation of steering wheel acceleration, reflect these well-known behaviors.

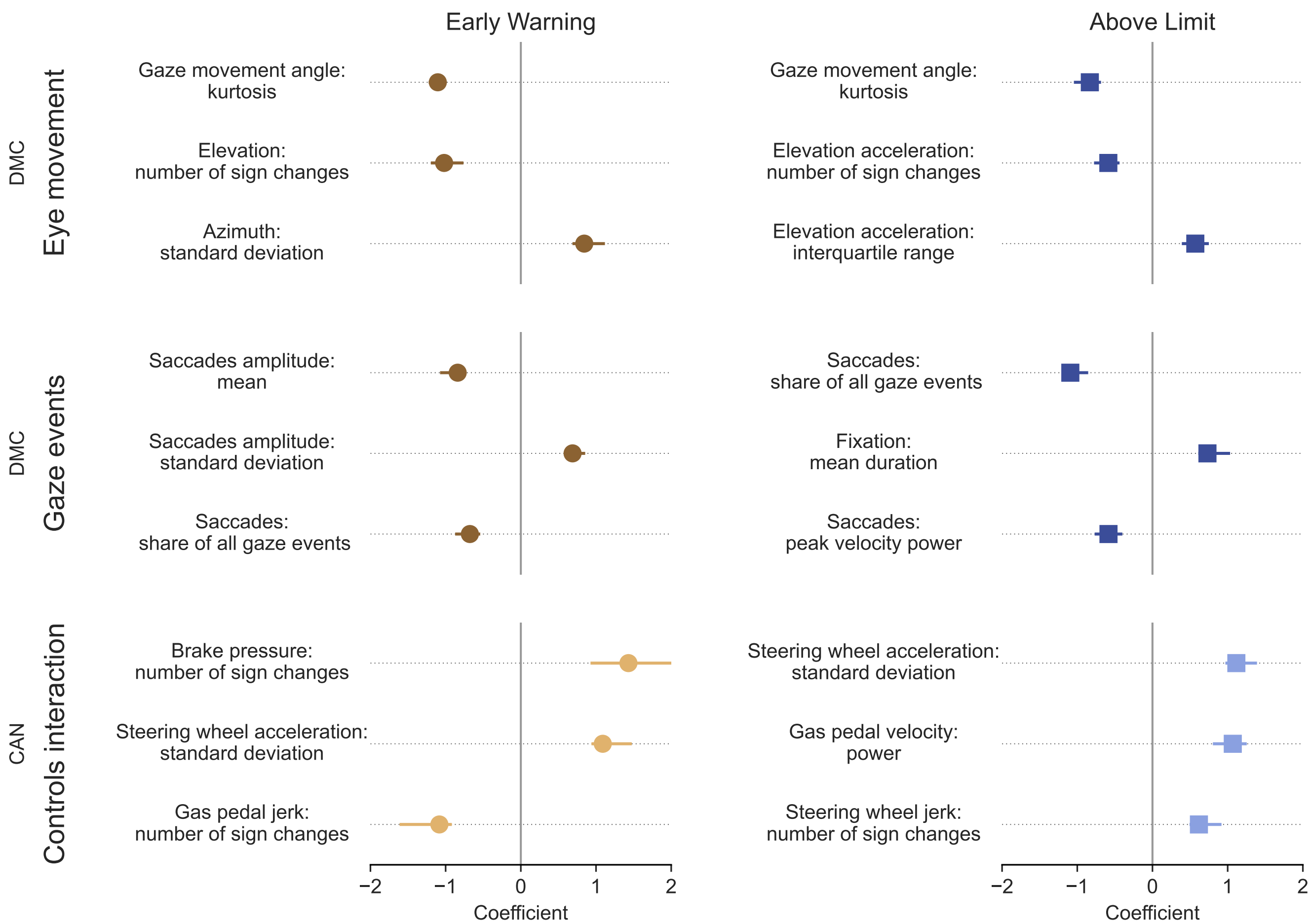

Figure 10: The figure highlights the three features with the largest absolute coefficients within each of the three most relevant feature groups. Error bars represent the minimum and maximum range of these coefficients obtained from cross-validation.

6.2 Comparison to previous work

The system we proposed achieved an AUROC of 0.84 ± 0.11 in classifying sobriety (Early Warning) and 0.80 ± 0.10 for Above Limit, utilizing both DMC and CAN data. With exclusive reliance on DMC data, the AUROC values recorded were 0.79 ± 0.12 for Early Warning and 0.75 ± 0.10 for Above Limit. Utilizing only CAN bus data, we noted an AUROC of 0.75 ± 0.13 for Early Warning and 0.72 ± 0.10 for Above Limit.

In a related study conducted by US NHTSA [59], researchers developed a CAN data-based classifier to predict whether a driver's BAC was above or below 0.08 g/dL, achieving an AUROC of 0.77 ± 0.08. Our study, which classifies at a lower threshold (0.05 g/dL) and incorporates out-of-sample validation, matched this performance with an AUROC of 0.72 ± 0.10 when using CAN bus data alone, improving further upon integrating DMC data.

Our objective was to adapt and apply the state-of-the-art approach of the original study to real vehicles, reproducing similar performance on the test track. Koch and Maritsch et al. achieved an AUROC of 0.88 ± 0.09 for Early Warning and 0.79 ± 0.10 for Above Limit using only DMC in a research-grade simulator with $n = 30$ participants driving under three BAC levels. When they combined DMC and CAN data, their system showed slightly enhanced AUROC scores of 0.91 ± 0.07 for Early Warning and 0.81 ± 0.11 for Above Limit.

The original study and our test track study showed high consistency in their detection performance. In line with the results of the original study, our confusion matrix revealed that the ML system demonstrated better performance for intoxication levels further away from the binary classification threshold. Specifically, for Early Warning, the FN rate for moderately intoxicated participants was 30%. This rate decreased to 17% for participants who were severely intoxicated. A corresponding pattern emerged for Above Limit predictions, with a FP rate of 36% for moderately intoxicated participants and a FP rate of 16% for sober participants. This finding is well-explainable, as greater differences in intoxication lead to more pronounced differences in symptoms [23], which makes the classification task easier.

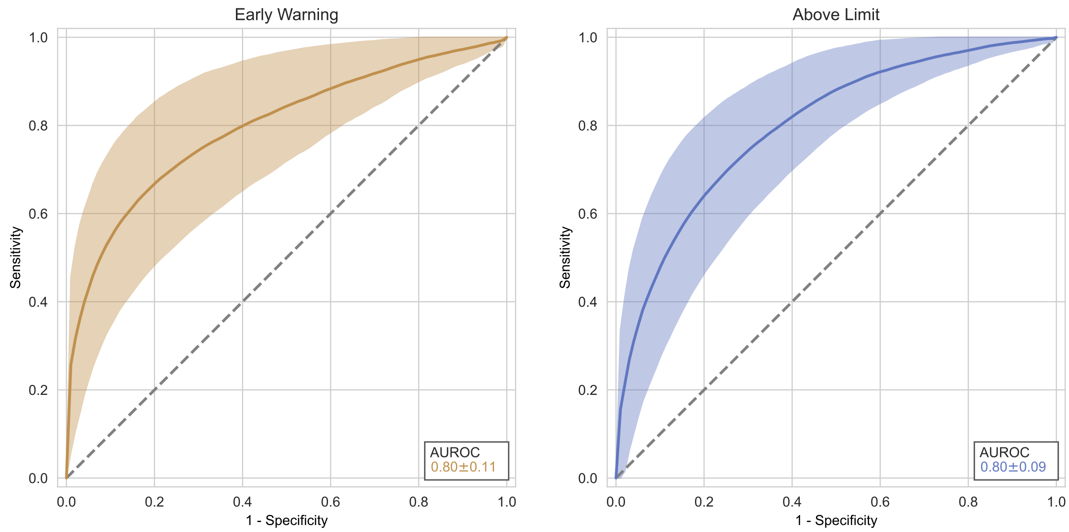

Figure 11: Performance with control groups: The figure displays subplots of receiver operating characteristic curves, averaged across all participants. The shaded areas represent the standard deviation among participants. The performance is reported as the area under the receiver operating characteristic curve (AUROC). A dashed gray line in each subplot represents the performance of a random classifier. Each column corresponds to a distinct classification task: Early Warning and Above Limit.

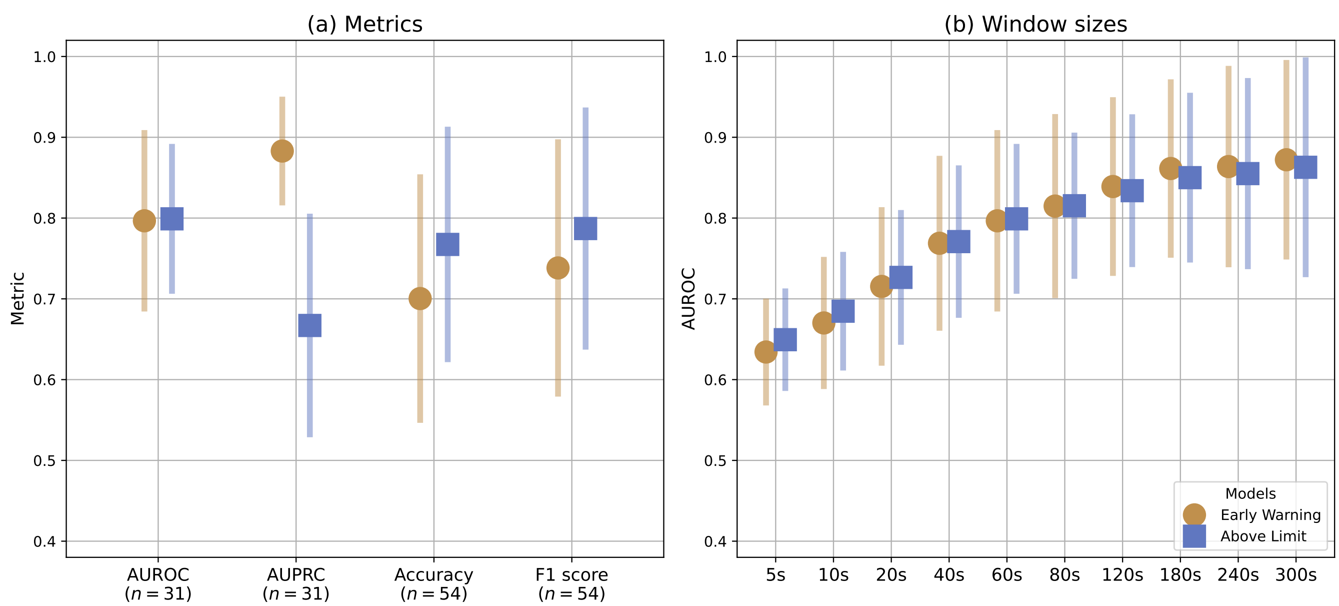

Figure 12: Performance with control groups: Subplot (a) illustrates the performance metrics area under the receiver operating characteristic curve (AUROC), area under the precision-recall curve (AUPRC), accuracy, and F1 score (weighted by class). Error bars represent the standard deviations. Each metric is evaluated for the two classification tasks' models: Early Warning and Above Limit. Subplot (b) details varying window sizes. The performance is quantified using AUROC. Error bars represent the standard deviations.

In literature, the generalizability of methodologies and results from simulator to the real vehicle remains a debated issue. Three primary factors potentially limit generalizability: the fidelity of the simulation environment [11, 19, 37], exclusion of real-world driving variables [61], and altered driver behavior due to no real risk [27]. Our use of industry-standard sensors, as opposed to high precision lab eye-tracking devices, enhanced the real-world applicability of our findings. We incorporated additional features such as region-specific gaze events to accommodate the wider field of view in actual vehicles, although some features like lateral lane positioning were inapplicable. By retaining the DMC-based features and incorporating CAN-based ones, we maintained performance parity with the original simulator study.

Moreover, we went beyond the original study by avoiding an exclusive within-subject design, instead employing a mixed design with both control and reference groups to bolster our study's robustness. Considering mixed evidence on placebo effects [34, 76] and the potential for model learning from driver training and drowsiness, our design aimed to mitigate these influences. Koch and Maritsch et al. addressed potential biases by implementing extensive simulator training and scheduling prolonged breaks to mitigate training and fatigue effects. We countered these potential biases by employing a reference control group that was fully informed (no placebo effect). Our study design emphasized the mitigation of learning biases from environmental changes, such as varying lighting conditions throughout the day. When incorporating the control groups, we successfully maintained consistent results, achieving an AUROC of 0.80 ± 0.11 for Early Warning and 0.80 ± 0.09 for Above Limit. This performance aligns with the mean AUROC scores observed in results without control groups.

In summary, we developed and evaluated a ML system aimed at detecting drunk driving using DMCs and real-time vehicle data. Given the real-world vehicle setting, which precluded the use of high-precision laboratory hardware, we adapted existing and introduced new features, incorporating real-time CAN vehicle data into our ML models.

6.3 Risk mitigation interventions

Upon predicting the driver's intoxication status, a personalized digital intervention (e.g., in-vehicle drunk driver warning) can be initiated to mitigate harmful drinking behaviors [48]. Furthermore, drivers who underestimate their BAC have been identified as particularly risky in previous studies [56]. For these individuals, an appropriate warning (such as auditory, visual, or tactile alerts) may effectively heighten awareness of their impairment.

In scenarios where a driver knowingly operates a vehicle with an excessive BAC, the efficacy of warnings may be inherently limited. By incorporating BAC levels into the driver behavior model, ADAS in the vehicle can adjust for the impaired motor functions and delayed reaction times of the driver. For instance, a collision prevention system might trigger a brake maneuver earlier knowing that the driver is highly likely to be intoxicated. As a last resort, the vehicle could be brought to a stop, although this measure requires further refinement given the current FP rates in detection.

6.4 Privacy and ethical considerations

Privacy discussions within the realm of information technology and surveillance (e.g., [67, 87]) concentrate on issues related to accessing private and personally identifiable data [75]. The concept of privacy is linked to various rights, including the right to be let alone, the right to control one's personal data, and the right to confidentiality [6]. More specifically, regulators worldwide have put forward a set of core privacy principles (e.g., Brazil [77], Singapore [80], and the EU [28]). Personal data must be processed lawfully, fairly, and transparently, ensuring individuals are informed about how their data is used. It should be collected for specific, legitimate purposes and not used in ways incompatible with those purposes. Data should be adequate, relevant, and limited to what is necessary, while being kept accurate and up-to-date. Personal data must not be stored longer than necessary. Finally, appropriate technical and organizational measures must protect data against unauthorized access, loss, or damage, ensuring its security and confidentiality [28].

Continuous driver monitoring involves processing highly sensitive data, thereby introducing substantial privacy risks. Notably, our detection approach does not rely on identifying a specific person. We have deliberately avoided personalizing the system, for example, by not deriving and applying individualized decision thresholds. In addition, we have minimized the need to store data. Specifically, we only need access to DMC and CAN data of the last 60 seconds. Furthermore, all calculations can be conducted onboard in a closed-loop system on the edge (in-vehicle). Committed to the principles of open science, we have published the source code of our system and made every effort to ensure that its functionality is easily understandable and transparent. Ultimately, however, vehicle manufactures have to assure that our approach is implemented in line with the outlined privacy principles. In regions like the EU where DMCs are mandated, data protection regulations in context of DMCs [29] require that systems for drowsiness and distraction detection retain data only as needed for their intended functions within a closed-loop system. This data must never be shared with or accessible to external parties and must be promptly erased post-processing [29]. These existing regulations are currently specific to drowsiness and distraction detection and hence need to be extended to cover impairment detection systems.

Beyond privacy concerns, the deployment of artificial intelligence and ML systems introduces several ethical challenges with significant societal implications [95]. For instance, biases in training data can result in unfair or discriminatory outcomes. Additionally, a lack of transparency in these systems might complicate efforts to understand or challenge their decisions, undermining accountability. To mitigate these issues, we have implemented interpretable ML techniques and ensured our system was trained on a high-quality data set. The diversity of our study participants was a key priority, as detailed in Table 2. Nonetheless, we recognize the need to enhance diversity further, such as by including a broader range of ethnicities, as discussed in Section 6.6.

6.5 Contributions

Harmful alcohol consumption and drunk driving pose significant public health challenges, contributing to disease, injuries, and fatalities. The advent of new vehicle generations equipped with DMCs presents opportunities to enhance road safety. To further mitigate accidents, active safety systems that comprehend and interact with the driver are essential. Given the advantages of integrating a driver behavior model into ADAS, we explored the efficacy of transferring drunk driving detection to the real vehicle, assessed the attainable performance levels, and identified the modifications needed for such implementation. Additionally, we examined the impact of including sober reference and placebo subjects on the performance metrics.

By adapting existing features (e.g., transitioning from gaze positions on a simulator display to three-dimensional gaze vectors), introducing new features (e.g., rotational head velocities), and incorporating new feature types, specifically region-specific gaze event features and eye state features, along with the inclusion of CAN-based real-time vehicle data, we conceptually replicated the simulator-based approach, with necessary modifications for the real-world setting. To the best of our knowledge, this work is the first to conduct a test track study as a clinical trial with intoxicated participants, utilizing DMC and CAN bus data to detect drunk driving. Our choice of model not only facilitated generalizability but also demonstrated robust performance on previously unseen drivers.

6.6 Limitations and future work

Given our data set size of $n = 54$, the variety of the collected data with respect to participants and driving behavior is naturally constrained. Despite this limitation, we endeavored to attract a diverse range of participants in terms of age and gender through targeted recruitment strategies. Table 2 illustrates that these efforts were successful with respect to the reported variables. Thus, we consider our participant set to be representative of the demographic composition of Switzerland. However, certain participant characteristics, such as ethnicity, were not recorded. Previous research indicates that eye-tracking performance can vary across different ethnic groups [7], suggesting potential areas for further improvement in future replications of the study with a broader variety and inclusion of special cases. Other factors may also influence driving and visual scanning behaviors. For instance, we hypothesize that our system might underperform with participants who have eye-related health issues. Moreover, studies have shown that the glance behaviors of older [24] and novice drivers [109] differ significantly from typical patterns, warranting additional investigation.

In line with the objectives of the original study, our replication study concentrates on exploring the general potential of DMC- and CAN-based drunk driving detection systems. We employ basic ML models and our goal was not to conduct rigorous comparisons between individual models. Consequently, future research should investigate more advanced methodologies and perform detailed analyses to discern significant performance differences across models.

The utilization of both, DMC data and CAN bus data enhances detection performance, but also increases hardware requirements and costs, which may limit applicability. However, the CAN bus is a standard feature in modern vehicles, while DMCs are mandatory in certain regions [29] and necessary for achieving high safety ratings [2, 33, 74]. Consequently, leveraging these standard vehicle systems assures cost-effectiveness. Although combining these modalities increases computational demand, our system does not require computationally intensive deep learning models frequently used in other studies for driver behavior analysis [94, 123]. In fact, the most resource-intensive component is extracting head and gaze features from the raw DMC images. The DMC operates at a frame rate of 50 Hz, while our drunk driving detection system assesses the driver's state at 1 Hz. Thus, eye, head, and gaze features are extracted more frequently than the frequency at which our ML system predicts the driver's state. However, in the automotive industry, real-time head and gaze detection systems that surpass our system's 50 Hz frequency are commercially available [21] and have become standard in new vehicles. Therefore, utilizing a combination of CAN bus and DMC data for drunk driving detection not only improves performance but is also practically viable.

The widespread adoption of fully autonomous vehicles in the coming decades is rather unlikely [20, 38, 71]. However, the integration of autonomous functionalities into the automotive market is progressing. Currently, Level 3 [89] autonomous cars (i.e., vehicles that can handle driving tasks within specific conditions, but require the driver to supervise) are commercially available, and forecasts indicate a growing prevalence of vehicles featuring autonomy. Consequently, the nature of driver-vehicle interactions is evolving. Drivers are increasingly disengaging from the driving process, which can diminish their situational awareness. The impact of these changes on the two sensor modalities we utilize, DMC and CAN, varies, especially in vehicles with autonomous features. Recent test track studies have analyzed the gaze behavior of drivers under the influence of alcohol (i.e., no drunk driving *detection*) in both, manual and autonomous modes. Zemblys et al. [129] showed that while high-level glance features (e.g., areas of interest) vary depending on the driving mode, basic gaze metrics such as fixations and saccades remain consistent across modes. Notably, fixation durations increased with intoxication irrespective of the mode. Tivesten et al. [106] concentrated on non-driving-related tasks across different driving modes, focusing on high-level glance features. Their findings also confirmed that intoxication affects gaze behavior regardless of the driving mode. However, with the ongoing advent of automation, the relevance of CAN features may diminish, potentially reducing the efficacy of CAN-based detection systems. In summary, our detection system, which utilizes both driver-environment interactions (DMC) and driver-vehicle interactions (CAN), is designed to function effectively even without direct driver-vehicle interactions, relying solely on DMC data. However, more work is necessary to rigorously validate our results in the context of autonomous vehicles.

While the validity of test track studies is generally high, it does not match the fidelity of real-world data [9]. During the test track study, we encountered environmental influences (e.g., severe rainfall, significant temperature fluctuations) typical of summer weather in Switzerland. However, our test track study, for example, did not involve other road users, non-driving related tasks, distractions, or nighttime driving conditions. In discussing the transferability from the simulator to test track studies, we identified three key simulator limitations: (1) varying degrees of realism [11, 19, 37], (2) the absence of external influences [61], and (3) modified driver behavior due to the minimized risk situation [27]. These limitations are also partially applicable to test track studies. Although our model demonstrated consistent performance across all test track driving scenarios, exploring these factors in real-world driving contexts, where driver's glance behavior is influenced more directly by the driving environment [107], would be valuable. Hence, while it entails higher risks and reduced experimental control, conducting studies in real-world contexts would provide higher validity [8]. Such data could be crucial for further validating our results and developing additional models. Consequently, we advocate for real-world studies to first validate our methods against FP in actual traffic conditions, and second, to generate a dataset featuring intoxicated drivers under real traffic conditions, naturally with legal exemptions, in alignment with local ethical standards, and with appropriate safety measures in place, such as a safety driver.

7 Conclusion

Society and policymakers are intensively exploring new strategies to curtail harmful alcohol consumption and eliminate drunk driving, both recognized for their fatal outcomes. For effective digital interventions, precise BAC estimation is imperative. Utilizing existing and emerging sensors offers a scalable and economical solution. Our test track study successfully replicated results from the state-of-the-art simulator study, affirming the generalizability of this approach. By employing human-interpretable features and a parsimonious model, we substantiated that the model is anchored in well-established physiological effects. Retaining DMC-based features and incorporating CAN-based ones, our system demonstrated a robust method for detecting drunk driving. Moreover, we introduced a novel study design with two control groups, effectively underlining the robustness of our system. Thus, our system sets the stage for pioneering digital interventions that aim to significantly reduce societal harm.

Acknowledgments

This work received funding and support from the Bosch Lab at the University of St. Gallen and ETH Zürich. The opinions and conclusions expressed in this paper are solely those of the authors and do not necessarily reflect the official policies, whether explicit or implicit, of the funding agencies. The sponsors did not influence the design, execution, data management, analysis, review, or interpretation of the research presented.

References

[1] Tousif Ahmed, Md Mahbubur Rahman, Ebrahim Nemati, Mohsin Yusuf Ahmed, Jilong Kuang, and Alex Jun Gao. 2023. Remote Breathing Rate Tracking in Stationary Position Using the Motion and Acoustic Sensors of Earables. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, USA, 1–22. https://doi.org/10.1145/3544548.3581265

[2] Australasian New Car Assessment Program (ANCAP). 2023. *How Is Car Safety Changing?* Australasian New Car Assessment Program (ANCAP). Retrieved August 28, 2024 from https://www.ancap.com.au/how-is-car-safety-changing

[3] Automobile Club of Switzerland. 2024. *ACS. Automotive Passion.* Automobile Club of Switzerland. Retrieved February 7, 2025 from https://www.acs.ch/en/

[4] Mohamed Bekkar, Hassiba Kheliouane Djemaa, and Taklit Akrouf Alitouche. 2013. Evaluation measures for models assessment over imbalanced data sets. *Journal Information Engineering Applications* 3, 10 (2013), 13 pages.

[5] Marcus Bendtsen, Katarina Åsberg, and Jim McCambridge. 2022. Effectiveness of a digital intervention versus alcohol information for online help-seekers in Sweden: a randomised controlled trial. *BMC Medicine* 20, 1 (May 2022), 176 pages. https://doi.org/10.1186/s12916-022-02374-5

[6] Colin J. Bennett and Charles D. Raab. 2017. *The Governance of Privacy: Policy instruments in global perspective*. Routledge, Milton Park, England, UK. https://doi.org/10.4324/9781315199269

[7] Pieter Blignaut and Daniël Wium. 2014. Eye-tracking data quality as affected by ethnicity and experimental design. *Behavior Research Methods* 46, 1 (March 2014), 67–80. https://doi.org/10.3758/s13428-013-0343-0

[8] Christian-Nils Boda. 2017. Driver interaction with vulnerable road users: Understanding and modelling driver behaviour for the design and evaluation of intelligent safety systems. https://publications.lib.chalmers.se/records/fulltext/250519/250519.pdf

[9] Christian-Nils Boda, Marco Dozza, Katarina Bohman, Prateek Thalya, Annika Larsson, and Nils Lubbe. 2018. Modelling how drivers respond to a bicyclist crossing their path at an intersection: How do test track and driving simulator compare? *Accident Analysis & Prevention* 111 (Feb. 2018), 238–250. https://doi.org/10.1016/j.aap.2017.11.032

[10] John Brick. 2006. Standardization of Alcohol Calculations in Research. *Alcoholism: Clinical and Experimental Research* 30, 8 (July 2006), 1276–1287. https://doi.org/10.1111/j.1530-0277.2006.00155.x

[11] Fanjun Bu, Stacey Li, David Goedicke, Mark Colley, Gyanendra Sharma, and Wendy Ju. 2024. Portobello: Extending Driving Simulation from the Lab to the Road. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3613904.3642341

[12] Vadim A. Butakov and Petros Ioannou. 2015. Personalized Driver/Vehicle Lane Change Models for ADAS. *IEEE Transactions on Vehicular Technology* 64, 10 (Oct. 2015), 4422–4431. https://doi.org/10.1109/tvt.2014.2369522

[13] Huiqin Chen and Lei Chen. 2017. Support Vector Machine Classification of Drunk Driving Behaviour. *International Journal of Environmental Research and Public Health* 14, 1 (Jan. 2017), 108. https://doi.org/10.3390/ijerph14010108

[14] Huiqin Chen, Xiuhong Yuan, Hongxian Ye, Lei Chen, and Guanjun Zhang. 2019. The effect of alcohol on the physiological performance of the driver. *International Journal of Crashworthiness* 24, 6 (2019), 656–663. https://doi.org/10.1080/13588265.2018.1511226

[15] Youngjun Cho. 2021. Rethinking Eye-blink: Assessing Task Difficulty through Physiological Representation of Spontaneous Blinking. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3411764.3445577

[16] Evangelia Christodoulou, Jie Ma, Gary S. Collins, Ewout W. Steyerberg, Jan Y. Verbakel, and Ben Van Calster. 2019. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology* 110 (June 2019), 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

[17] Asim H. Dar, Adina S. Wagner, and Michael Hanke. 2021. REMoDNaV: robust eye-movement classification for dynamic stimulation. *Behavior Research Methods* 53, 1 (Feb. 2021), 399–414. https://doi.org/10.3758/s13428-020-01428-x

[18] Ksander N. de Winkel, Tugrul Irmak, Riender Happee, and Barys Shyrokau. 2023. Standards for passenger comfort in automated vehicles: Acceleration and jerk. *Applied Ergonomics* 106 (Jan. 2023), 9 pages. https://doi.org/10.1016/j.apergo.2022.103881

[19] Joost De Winter, Peter M van Leeuwen, Riender Happee, et al. 2012. Advantages and Disadvantages of Driving Simulators: A Discussion. In *Proceedings of Measuring Behavior*, Vol. 2012. Noldus, Utrecht, Netherlands, 47–50.

[20] Johannes Deichmann, Eike Ebel, Kersten Heineke, Ruth Heuss, and Martin Kellner. 2023. *Autonomous driving's future: Convenient and connected.* McKinsey & Company. Retrieved August 29, 2024 from https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/autonomous-drivings-future-convenient-and-connected#/

[21] Yanchao Dong, Zhencheng Hu, Keiichi Uchimura, and Nobuki Murayama. 2009. Driver inattention monitoring system for intelligent vehicles: A review, In 2009 IEEE Intelligent Vehicles Symposium. *IEEE transactions on intelligent transportation systems* 12, 2, 875–880. https://doi.org/10.1109/ivs.2009.5164395

[22] Messaoud Doudou, Abdelmadjid Bouabdallah, and Véronique Berge-Cherfaoui. 2020. Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges. *International Journal of Intelligent Transportation Systems Research* 18, 2 (Sept. 2020), 297–319. https://doi.org/10.1007/s13177-019-00199-w

[23] Matthew J. Dry, Nicholas R. Burns, Ted Nettelbeck, Aaron L. Farquharson, and Jason M. White. 2012. Dose-Related Effects of Alcohol on Cognitive Functioning. *PloS one* 7, 11 (Nov. 2012), e50977. https://doi.org/10.1371/journal.pone.0050977

[24] Tania Dukic and Thomas Broberg. 2012. Older drivers' visual search behaviour at intersections. *Transportation Research Part F: Traffic Psychology and Behaviour* 15, 4 (July 2012), 462–470. https://doi.org/10.1016/j.trf.2011.10.001

[25] Florian Echtler and Maximilian Häußler. 2018. Open Source, Open Science, and the Replication Crisis in HCI. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/3170427.3188395

[26] Eidgenössisches Justiz- und Polizeidepartement (EJPD). 2016. Verordnung des EJPD über Atemalkoholmessmittel (AAMV). https://www.fedlex.admin.ch/eli/cc/2015/88/de

[27] Stéphane Espié, Pierre Gauriat, and Max Duraz. 2005. Driving Simulators Validation: The Issue of Transferability of Results Acquired on Simulator. In *Driving Simulation Conference North-America (DSC-NA 2005), Orlondo, FL*. Center for Advanced Transportation Systems Simulation, Orlando, FL, USA, 149–156. https://www.nads-sc.uiowa.edu/dscna/2005/main.htm

[28] European Parliament and Council of the European Union. 2016. Regulation (EU) 2016/679. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679

[29] European Parliament and Council of the European Union. 2019. Regulation (EU) 2019/2144. http://data.europa.eu/eli/reg/2019/2144/oj

[30] Rosa L Figueroa, Qing Zeng-Treitler, Sasikiran Kandula, and Long H Ngo. 2012. Predicting sample size required for classification performance. *BMC Medical Informatics and Decision Making* 12, 1 (Feb. 2012), 1–10. https://doi.org/10.1186/1472-6947-12-8

[31] James M. Fleming, Craig K. Allison, Xingda Yan, Roberto Lot, and Neville A. Stanton. 2019. Adaptive driver modelling in ADAS to improve user acceptance: A study using naturalistic data. *Safety Science* 119 (Nov. 2019), 76–83. https://doi.org/10.1016/j.ssci.2018.08.023

[32] Per-Anders Fransson, Fredrik Modig, Mitesh Patel, Stephen Gomez, and Måns Magnusson. 2010. Oculomotor deficits caused by 0.06% and 0.10% blood alcohol concentrations and relationship to subjective perception of drunkenness. *Clinical Neurophysiology* 121, 12 (Dec. 2010), 2134–2142. https://doi.org/10.1016/j.clinph.2010.05.003

[33] Rikard Fredriksson, Michael G. Lenné, Sjef van Montfort, and Colin Grover. 2021. European NCAP Program Developments to Address Driver Distraction, Drowsiness and Sudden Sickness. *Frontiers in Neuroergonomics* 2 (Nov. 2021), 33 pages. https://doi.org/10.3389/fnrgo.2021.786674

[34] María Nerea Galindo, José Francisco Navarro, and María Cavas. 2020. The Influence of Placebo Effect on Craving and Cognitive Performance in Alcohol, Caffeine, or Nicotine Consumers: A Systematic Review. *Frontiers in Psychiatry* 11 (Aug. 2020), 849. https://doi.org/10.3389/fpsyt.2020.00849

[35] Davoud Gholamiangonabadi, Nikita Kiselov, and Katarina Grolinger. 2020. Deep Neural Networks for Human Activity Recognition With Wearable Sensors: Leave-One-Subject-Out Cross-Validation for Model Selection. *IEEE Access* 8 (2020), 133982–133994. https://doi.org/10.1109/access.2020.3010715

[36] Rebecca Gilbertson, Robert Prather, and Sara Jo Nixon. 2010. Acute Alcohol Administration and Placebo Effectiveness in Older Moderate Drinkers: Influences on Cognitive Performance. *Journal of Studies on Alcohol and Drugs* 71, 3 (May 2010), 345–350. https://doi.org/10.15288/jsad.2010.71.345

[37] David Goedicke, Alexandra WD Bremers, Sam Lee, Fanjun Bu, Hiroshi Yasuda, and Wendy Ju. 2022. XR-OOM: MiXed Reality driving simulation with real cars for research and design. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3491102.3517704

[38] Goldman Sachs. 2024. *Partially autonomous cars forecast to comprise 10% of new vehicle sales by 2030*. Goldman Sachs. Retrieved August 20, 2024 from https://www.goldmansachs.com/insights/articles/partially-autonomous-cars-forecast-to-comprise-10-percent-of-new-vehicle-sales-by-2030#

[39] Jing-Ming Guo and Herleeyandi Markoni. 2019. Driver drowsiness detection using hybrid convolutional neural network and long short-term memory. *Multimedia Tools and Applications* 78, 20 (Oct. 2019), 29059–29087. https://doi.org/10.1007/s11042-018-6378-6

[40] David H. Gustafson, Fiona M. McTavish, Ming-Yuan Chih, Amy K. Atwood, Roberta A. Johnson, Michael G. Boyle, Michael S. Levy, Hilary Driscoll, Steven M. Chisholm, Lisa Dillenburg, Andrew Isham, and Dhavan Shah. 2014. A Smartphone Application to Support Recovery From Alcoholism: A Randomized Clinical Trial. *JAMA Psychiatry* 71, 5 (May 2014), 566–572. https://doi.org/10.1001/jamapsychiatry.2013.4642

[41] Roland Gustafson. 1988. Effects of alcohol on power in social interaction between man and woman. *Journal of Studies on Alcohol* 49, 1 (Jan. 1988), 78–84. https://doi.org/10.15288/jsa.1988.49.78

[42] Hasanin Harkous and Hassan Artail. 2019. A Two-Stage Machine Learning Method for Highly-Accurate Drunk Driving Detection. In *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, IEEE, New York, NY, USA, 1–6. https://doi.org/10.1109/wimob.2019.8923366

[43] Hasanin Harkous, Carine Bardawil, Hassan Artail, and Naseem Daher. 2018. Application of Hidden Markov Model on Car Sensors for Detecting Drunk Drivers. In *2018 IEEE International Multidisciplinary Conference on Engineering Technology (IMCET)*. IEEE, IEEE, New York, NY, USA, 1–6. https://doi.org/10.1109/imcet.2018.8603030

[44] Anuradha Herath, Bradley Rey, Sandra Bardot, Sawyer Rempel, Lucas Audette, Huizhe Zheng, Jun Li, Kevin Fan, Da-Yuan Huang, Wei Li, and Pourang Irani. 2022. Expanding Touch Interaction Capabilities for Smart-rings: An Exploration of Continual Slide and Microroll Gestures. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '22)*. ACM, New York, NY, USA, 1–7. https://doi.org/10.1145/3491101.3519714

[45] Quanan Huang and Huiyi Wang. 2004. Fundamental Study of Jerk: Evaluation of Shift Quality and Ride Comfort. In *SAE Technical Paper Series (ADSC)*. SAE International, Warrendale, PA, USA, Article 2065, 8 pages. https://doi.org/10.4271/2004-01-2065

[46] Christopher Irwin, Elizaveta Iudakhina, Ben Desbrow, and Danielle McCartney. 2017. Effects of acute alcohol consumption on measures of simulated driving: a systematic review and meta-analysis. *Accident Analysis & Prevention* 102 (May 2017), 248–266. https://doi.org/10.1016/j.aap.2017.03.001

[47] S. Jongen, E.F.P.M. Vuurman, J.G. Ramaekers, and A. Vermeeren. 2016. The sensitivity of laboratory tests assessing driving related skills to dose-related impairment of alcohol: A literature review. *Accident Analysis & Prevention* 89 (April 2016), 31–48. https://doi.org/10.1016/j.aap.2016.01.001

[48] Eileen FS Kaner, Fiona R Beyer, Claire Garnett, David Crane, Jamie Brown, Colin Muirhead, James Redmore, Amy O'Donnell, James J Newham, Frank de Vocht, Matthew Hickman, Heather Brown, Gregory Maniatopoulos, and Susan Michie. 2017. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. *Cochrane Database of Systematic Reviews* 2017, 9, Article CD011479 (Sept. 2017), 157 pages. https://doi.org/10.1002/14651858.cd011479.pub2

[49] Alexey Kashevnik, Roman Shchedrin, Christian Kaiser, and Alexander Stocker. 2021. Driver Distraction Detection Methods: A Literature Review and Framework. *IEEE Access* 9 (2021), 60063–60076. https://doi.org/10.1109/access.2021.3073599

[50] Terence M. Keane, Stephen A. Lisman, and Jeffrey Kreutzer. 1980. Alcoholic beverages and their placebos: An empirical evaluation of expectancies. *Addictive Behaviors* 5, 4 (Jan. 1980), 313–328. https://doi.org/10.1016/0306-4603(80)90005-2

[51] Timothy Kerns. 2017. *Effectiveness of an ignition interlock device in reducing alcohol-impaired driving Recidivism and Alcohol-Impaired Motor Vehicle Crashes in Maryland*. Ph. D. Dissertation. University of Maryland.

[52] Ensiyeh Keshtkaran, Brodie von Berg, Grant Regan, David Suter, and Syed Zulqarnain Gilani. 2024. Estimating Blood Alcohol Level Through Facial Features for Driver Impairment Assessment. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. IEEE, New York, NY, USA, 4527–4536. https://doi.org/10.1109/wacv57701.2024.00448

[53] Katja Kircher, Svitlana Finér, Raimondas Zemblys, and Christer Ahlström. 2023. *Fit2Drive*. Technical Report. Fordonsstrategisk Forskning och Innovation, Sweden.

[54] Kevin Koch, Martin Maritsch, Eva Van Weenen, Stefan Feuerriegel, Matthias Pfäffli, Elgar Fleisch, Wolfgang Weinmann, and Felix Wortmann. 2023. Leveraging driver vehicle and environment interaction: Machine learning using driver monitoring cameras to detect drunk driving. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. ACM, New York, NY, USA, 1–32. https://doi.org/10.1145/3544548.3580975

[55] Grégoire S. Larue, Christian Wullems, Michelle Sheldrake, and Andry Rakotonirainy. 2018. Validation of a Driving Simulator Study on Driver Behavior at Passive Rail Level Crossings. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 60, 6 (July 2018), 743–754. https://doi.org/10.1177/0018720818783507

[56] Jennifer R. Laude and Mark T. Fillmore. 2016. Drivers who self-estimate lower blood alcohol concentrations are riskier drivers after drinking. *Psychopharmacology* 233, 8 (Feb. 2016), 1387–1394. https://doi.org/10.1007/s00213-016-4233-x

[57] Samuel Lawoyin, Ding-Yu Fei, and Ou Bai. 2015. Accelerometer-based steering-wheel movement monitoring for drowsy-driving detection. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* 229, 2 (2015), 163–173. https://doi.org/10.1177/0954407014536148

[58] Hansoo Lee, Auk Kim, SangWon Bae, and Uichin Lee. 2024. S-ADL: Exploring Smartphone-based Activities of Daily Living to Detect Blood Alcohol Concentration in a Controlled Environment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. ACM, New York, NY, USA, 1–25. https://doi.org/10.1145/3613904.3642832

[59] John D. Lee, Dary Fiorentino, Michelle L. Reyes, Timothy L. Brown, Omar Ahmad, James Fell, Nic Ward, and Robert Dufour. 2010. *Assessing the Feasibility of Vehicle-Based Sensors to Detect Alcohol Impairment*. Technical Report. National Highway Traffic Safety Administration, Washington, DC, USA. 98 pages.

[60] Kang Hee Lee, Keon Hee Baek, Su Bin Choi, Nak Tak Jeong, Hyung Uk Moon, Eun Seong Lee, Hyung Min Kim, and Myung Won Suh. 2019. Development of Three Driver State Detection Models from Driving Information Using Vehicle Simulator; Normal, Drowsy and Drunk Driving. *International Journal of Automotive Technology* 20, 6 (Sept. 2019), 1205–1219. https://doi.org/10.1007/s12239-019-0113-9

[61] Vera Lehmann, Thomas Zueger, Martin Maritsch, Michael Notter, Simon Schallmoser, Caterina Bérubé, Caroline Albrecht, Mathias Kraus, Stefan Feuerriegel, Elgar Fleisch, Tobias Kowatsch, Sophie Lagger, Markus Laimer, Felix Wortmann, and Christoph Stettler. 2024. Machine Learning to Infer a Health State

Using Biomedical Signals — Detection of Hypoglycemia in People with Diabetes while Driving Real Cars. *NEJM AI* 1, 3 (Feb. 2024), 10 pages. https://doi.org/10.1056/aioa2300013

[62] Zhenlong Li, Xue Jin, and Xiaohua Zhao. 2015. Drunk driving detection based on classification of multivariate time series. *Journal of Safety Research* 54 (Sept. 2015), 61.e29–64. https://doi.org/10.1016/j.jsr.2015.06.007

[63] ZhenLong Li, HaoXin Wang, YaoWei Zhang, and XiaoHua Zhao. 2020. Random forest–based feature selection and detection method for drunk driving recognition. *International Journal of Distributed Sensor Networks* 16, 2 (Feb. 2020), 13 pages. https://doi.org/10.1177/1550147720905234

[64] Lars Lindberg, David Grubb, Daniel Dencker, Mikael Finnhult, and Sven-Gunnar Olsson. 2015. Detection of mouth alcohol during breath alcohol analysis. *Forensic Science International* 249 (April 2015), 66–72. https://doi.org/10.1016/j.forsciint.2015.01.017

[65] Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions. *ACM Transactions on Management Information Systems (TMIS)* 4, 4 (Dec. 2013), 1–28. https://doi.org/10.1145/2544103

[66] Marc Luginbühl, Friedrich M. Wurst, Frederike Stöth, Wolfgang Weinmann, Christophe P. Stove, and Katleen Van Uytfanghe. 2022. Consensus for the use of the alcohol biomarker phosphatidylethanol (PEth) for the assessment of abstinence and alcohol consumption in clinical and forensic practice (2022 Consensus of Basel). *Drug Testing and Analysis* 14, 10 (July 2022), 1800–1802. https://doi.org/10.1002/dta.3340

[67] Kevin Macnish. 2017. *The Ethics of Surveillance: An Introduction.* Routledge, Milton Park, England, UK. https://doi.org/10.4324/9781315162867

[68] Alex Mariakakis, Sayna Parsi, Shwetak N. Patel, and Jacob O. Wobbrock. 2018. Drunk User Interfaces: Determining Blood Alcohol Level through Everyday Smartphone Tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3173808

[69] Pierre Maurage, Nicolas Masson, Zoé Bollen, and Fabien D'Hondt. 2020. Eye tracking correlates of acute alcohol consumption: A systematic and critical review. *Neuroscience & Biobehavioral Reviews* 108 (Jan. 2020), 400–422. https://doi.org/10.1016/j.neubiorev.2019.10.001

[70] Lakmal Meegahapola, Florian Labhart, Thanh-Trung Phan, and Daniel Gatica-Perez. 2021. Examining the Social Context of Alcohol Drinking in Young Adults with Smartphone Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (Sept. 2021), 1–26. https://doi.org/10.1145/3478126

[71] Dimitris Milakis, Maaike Snelder, Bart Van Arem, Bert Van Wee, and Gonçalo Homem de Almeida Correia. 2017. Development and transport implications of automated vehicles in the Netherlands: scenarios for 2030 and 2050. *European Journal of Transport and Infrastructure Research* 17, 1 (2017), 63–85. https://doi.org/10.18757/ejtir.2017.17.1.3180

[72] A. Moser, Wolfgang Heide, and Detlef Kömpf. 1998. The effect of oral ethanol consumption on eye movements in healthy volunteers. *Journal of Neurology* 245, 8 (Aug. 1998), 542–550. https://doi.org/10.1007/s004150050240

[73] Herbert Moskowitz and Dary Fiorentino. 2000. *A Review of the Literature on the Effects of Low Doses of Alcohol on Driving-Related Skills.* Technical Report. National Highway Traffic Safety Administration, Washington, DC, USA. https://doi.org/10.21949/1525468

[74] Megan Mulhall, Kyle Wilson, Shiyan Yang, Jonny Kuo, Tracey Sletten, Clare Anderson, Mark E. Howard, Shantha Rajaratnam, Michelle Magee, Allison Collins, and Michael G. Lenné. 2023. European NCAP Driver State Monitoring Protocols: Prevalence of Distraction in Naturalistic Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 66, 9 (Aug. 2023), 2205–2217. https://doi.org/10.1177/00187208231194543

[75] Vincent C. Müller. 2023. Ethics of Artificial Intelligence and Robotics. In *The Stanford Encyclopedia of Philosophy* (fall 2023 ed.), Edward N. Zalta and Uri Nodelman (Eds.). Metaphysics Research Lab, Stanford University, Stanford, CA, USA.

[76] Craig T. Nagoshi, Richard T. Noll, and Mark D. Wood. 1992. Alcohol Expectancies and Behavioral and Emotional Responses to Placebo Versus Alcohol Administration. *Alcoholism: Clinical and Experimental Research* 16, 2 (April 1992), 255–260. https://doi.org/10.1111/j.1530-0277.1992.tb01372.x

[77] National Congress of Brazil. 2018. Lei Geral de Proteção de Dados Pessoais (LGPD), Lei nº 13.709. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm

[78] National Highway Traffic Safety Administration. 2023. *Advanced Impaired Driving Prevention Technology.* Technical Report. National Highway Traffic Safety Administration, Washington, DC, USA.

[79] Sigrid Nyberg, Göran Wahlström, Torbjörn Bäckström, and Inger Sundström Poromaa. 2004. No difference in responsiveness to a low dose of alcohol between healthy women and men. *Pharmacology Biochemistry and Behavior* 78, 3 (July 2004), 603–610. https://doi.org/10.1016/j.pbb.2004.03.027

[80] Parliament of Singapore. 2012. Personal Data Protection Act 2012, Act 26 of 2012. https://sso.agc.gov.sg/Act/PDPA2012

[81] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.

[82] Ignacio Perez-Pozuelo, Dimitris Spathis, Emma A.D. Clifton, and Cecilia Mascolo. 2021. *Wearables, smartphones, and artificial intelligence for digital phenotyping and health.* Elsevier, Amsterdam, Netherlands, 33–54. https://doi.org/10.1016/b978-0-12-820077-3.00003-1

[83] Céline Poisson, Jessy Barré, Gaëtan Bourmaud, and Jean-François Forzy. 2020. Driver Behavior in Conditional Automation: Comparison of Driving Simulator and Wizard of Oz Conditions. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20).* ACM, New York, NY, USA, 1–7. https://doi.org/10.1145/3334480.3382854

[84] Johannes Gerardus Ramaekers, Hindrik W. J. Robbe, and J.F. O'Hanlon. 2000. Marijuana, alcohol and actual driving performance. *Human Psychopharmacology: Clinical and Experimental* 15, 7 (2000), 551–558. https://doi.org/10.1002/1099-1077(200010)15:7<551::aid-hup236>3.0.co;2-p

[85] Muhammad Ramzan, Hikmat Ullah Khan, Shahid Mahmood Awan, Amina Ismail, Mahwish Ilyas, and Ahsan Mahmood. 2019. A Survey on State-of-the-Art Drowsiness Detection Techniques. *IEEE Access* 7 (2019), 61904–61919. https://doi.org/10.1109/access.2019.2914373

[86] Daniel J. O. Roche and Andrea C. King. 2010. Alcohol impairment of saccadic and smooth pursuit eye movements: impact of risk factors for alcohol dependence. *Psychopharmacology* 212, 1 (July 2010), 33–44. https://doi.org/10.1007/s00213-010-1906-8

[87] Beate Roessler. 2017. X—Privacy as a Human Right. *Proceedings of the Aristotelian Society* 117, 2 (July 2017), 187–206. https://doi.org/10.1093/arisoc/aox008

[88] Ann M. Romosz, Michael Scherer, Robert B. Voas, Eduardo Romano, Thomas H. Nochajski, Eileen P. Taylor, Meg R. Brogdale, and Amy R. Manning. 2021. Understanding non-installers of the ignition interlock device: A qualitative analysis. *Drug and Alcohol Review* 40, 6 (March 2021), 1083–1091. https://doi.org/10.1111/dar.13275

[89] SAE International. 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. https://doi.org/10.4271/J3016_202104

[90] John B Saunders, Olaf G Aasland, Thomas F Babor, Juan R De la Fuente, and Marcus Grant. 1993. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. *Addiction* 88, 6 (June 1993), 791–804. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

[91] Michael A. Sayette, F.Curtis Breslin, G.Terence Wilson, and Gianine D. Rosenblum. 1994. An evaluation of the balanced placebo design in alcohol administration research. *Addictive Behaviors* 19, 3 (May 1994), 333–342. https://doi.org/10.1016/0306-4603(94)90034-5

[92] Kai-Uwe Schmitt, Christian Lanz, Markus H. Muser, Felix Walz, and Urs Schwarz. 2013. Saccadic eye movements after low-dose oral alcohol exposure. *Journal of Forensic and Legal Medicine* 20, 7 (Oct. 2013), 870–874. https://doi.org/10.1016/j.jflm.2013.06.023

[93] Alexandra Schröck, Annette Thierauf-Emberger, Stefan Schürch, and Wolfgang Weinmann. 2017. Phosphatidylethanol (PEth) detected in blood for 3 to 12 days after single consumption of alcohol—a drinking study with 16 volunteers. *International Journal of Legal Medicine* 131, 1 (Sept. 2017), 153–160. https://doi.org/10.1007/s00414-016-1445-x

[94] Mohammad Shahverdy, Mahmood Fathy, Reza Berangi, and Mohammad Sabokrou. 2020. Driver behavior detection and classification using deep convolutional neural networks. *Expert Systems with Applications* 149 (July 2020), 12 pages. https://doi.org/10.1016/j.eswa.2020.113240

[95] Keng Siau and Weiyu Wang. 2020. Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI. *Journal of Database Management* 31, 2 (April 2020), 74–87. https://doi.org/10.4018/jdm.2020040105

[96] Jéssica Bruna Santana Silva, Eva Dias Cristino, Natalia Leandro de Almeida, Paloma Cavalcante Bezerra de Medeiros, and Natanael Antonio dos Santos. 2017. Effects of acute alcohol ingestion on eye movements and cognition: a double-blind, placebo-controlled study. *PLoS One* 12, 10 (Oct. 2017), 16 pages. https://doi.org/10.1371/journal.pone.0186061

[97] Smart Eye AB. 2023. *Smart Eye's Market-Leading Driver Monitoring Software Included in New Volvo EX90.* Smart Eye AB. Retrieved February 7, 2025 from https://www.smarteye.se/news/smart-eyes-market-leading-driver-monitoring-software-included-in-new-volvo-ex90/

[98] Ivneet Sohi, Kevin D Shield, Jürgen Rehm, and Maristela Monteiro. 2023. Digital interventions for reducing alcohol use in general populations: An updated systematic review and meta-analysis. *Alcohol: Clinical and Experimental Research* 47, 10 (Oct. 2023), 1813–1832. https://doi.org/10.1111/acer.15175

[99] Steve Stewart-Williams and John Podd. 2004. The Placebo Effect: Dissolving the Expectancy Versus Conditioning Debate. *Psychological Bulletin* 130, 2 (2004),

324–340. https://doi.org/10.1037/0033-2909.130.2.324

[100] Frederike Stöth, Kevin Koch, Matthias Bantle, Philipp Pütz, Felix Wortmann, and Wolfgang Weinmann. 2023. Increase of PEth after single consumption of alcohol and evaluation of a volumetric DBS filter paper device. *Journal of Analytical Toxicology* 47, 6 (May 2023), e51–e51. https://doi.org/10.1093/jat/bkad029

[101] David L. Strayer, Frank A. Drews, and Dennis J. Crouch. 2006. A Comparison of the Cell Phone Driver and the Drunk Driver. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48, 2 (June 2006), 381–391. https://doi.org/10.1518/001872006777724471

[102] Yifan Sun, Jinglei Zhang, Xiaoyuan Wang, Zhangu Wang, and Jie Yu. 2018. Recognition Method of Drinking-driving Behaviors Based on PCA and RBF Neural Network. *Promet-Traffic&Transportation* 30, 4 (Aug. 2018), 407–417. https://doi.org/10.7307/ptt.v30i4.2657

[103] Swiss Federal Institute of Metrology METAS. 2024. *Certificate Search*. Swiss Federal Institute of Metrology METAS. Retrieved February 7, 2025 from https://legnet.metas.ch/legnet2/Eichstellen/certsearch

[104] Swiss Federal Office of Topography swisstopo. 2024. *Map View*. Swiss Federal Office of Topography swisstopo. Retrieved February 7, 2025 from https://map.geo.admin.ch/

[105] Maria Testa, Mark T. Fillmore, Jeanette Norris, Antonia Abbey, John J. Curtin, Kenneth E. Leonard, Kristin A. Mariano, Margaret C. Thomas, Kim J. Nomensen, William H. George, Carol VanZile-Tamsen, Jennifer A. Livingston, Christopher Saenz, Philip O. Buck, Tina Zawacki, Michele R. Parkhill, Angela J. Jacques, and Lenwood W. Hayman. 2006. Understanding Alcohol Expectancy Effects: Revisiting the Placebo Condition. *Alcoholism: Clinical and Experimental Research* 30, 2 (Jan. 2006), 339–348. https://doi.org/10.1111/j.1530-0277.2006.00039.x

[106] Emma Tivesten, Viktor Broo, and Mikael Ljung Aust. 2023. The influence of alcohol and automation on drivers' visual behavior during test track driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 95 (May 2023), 215–227. https://doi.org/10.1016/j.trf.2023.04.008

[107] Emma Tivesten and Marco Dozza. 2014. Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 26 (Sept. 2014), 258–272. https://doi.org/10.1016/j.trf.2014.08.004

[108] Toyota. 2025. *T-Mate Driving Assistance - Your T-Mate while driving*. Toyota. Retrieved February 7, 2025 from https://www.toyota-europe.com/brands-and-services/toyota/t-mate-driving-assistance

[109] Geoffrey Underwood, Peter Chapman, Neil Brocklehurst, Jean Underwood, and David Crundall. 2003. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics* 46, 6 (Jan. 2003), 629–646. https://doi.org/10.1080/0014013031000090116

[110] United Nations (UN) General Assembly. 2015. Transforming our world: the 2030 Agenda for Sustainable Development. https://sdgs.un.org/2030agenda A/RES/70/1.

[111] United States Congress (117th). 2021. H.R.3684 - Infrastructure Investment and Jobs Act. https://www.congress.gov/bill/117th-congress/house-bill/3684/text

[112] Luuk Vissers, Sjoerd Houwing, and Fred Wegman. 2017. *Alcohol-Related Road Casualties in Official Crash Statistics*. Technical Report. OECD / ITF, Paris, France.

[113] Chat Wacharamanotham, Lukas Eisenring, Steve Haroz, and Florian Echtler. 2020. Transparency of CHI Research Artifacts: Results of a Self-Reported Survey. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376448

[114] Weichen Wang, Shayan Mirjafari, Gabriella Harari, Dror Ben-Zeev, Rachel Brian, Tanzeem Choudhury, Marta Hauser, John Kane, Kizito Masaba, Subigya Nepal, Akane Sano, Emily Scherer, Vincent Tseng, Rui Wang, Hongyi Wen, Jialing Wu, and Andrew Campbell. 2020. Social Sensing: Assessing Social Functioning of Patients Living with Schizophrenia using Mobile Phone Sensing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376855

[115] Patricia E. Watson, Ian D. Watson, and Richard D. Batt. 1981. Prediction of blood alcohol concentrations in human subjects. Updating the Widmark Equation. *Journal of Studies on Alcohol* 42, 7 (July 1981), 547–556. https://doi.org/10.15288/jsa.1981.42.547

[116] Erik Matteo Prochet Widmark. 1932. *Die theoretischen Grundlagen und die praktische Verwendbarkeit der gerichtlich-medizinischen Alkoholbestimmung*. Number 11 in Fortschritte der naturwissenschaftlichen Forschung. Berlin : Urban & Schwarzenberg, Berlin, Germany.

[117] Max Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, and Jeffrey Nichols. 2012. RepliCHI SIG: From a panel to a new submission venue for replication. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1185–1188. https://doi.org/10.1145/2212776.2212419

[118] Max L. Wilson, Ed H. Chi, Stuart Reeves, and David Coyle. 2014. RepliCHI: the workshop II. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 33–36. https://doi.org/10.1145/2559206.2559233

[119] Max L. Wilson, Wendy Mackay, Ed Chi, Michael Bernstein, Dan Russell, and Harold Thimbleby. 2011. RepliCHI-CHI should be replicating and validating results more: discuss. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 463–466. https://doi.org/10.1145/1979742.1979491

[120] Max L. L. Wilson, Paul Resnick, David Coyle, and Ed H. Chi. 2013. Replichi: the workshop. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 3159–3162. https://doi.org/10.1145/2468356.2479636

[121] World Health Organization (WHO). 2019. *The SAFER technical package: five areas of intervention at national and subnational levels*. Technical Report. World Health Organization (WHO), Geneva, Switzerland.

[122] World Health Organization (WHO). 2024. *Global status report on alcohol and health and treatment of substance use disorders*. Technical Report. World Health Organization (WHO), Geneva, Switzerland.

[123] Yang Xing, Chen Lv, Huaji Wang, Dongpu Cao, Efstathios Velenis, and Fei-Yue Wang. 2019. Driver Activity Recognition for Intelligent Vehicles: A Deep Learning Approach. *IEEE transactions on Vehicular Technology* 68, 6 (June 2019), 5379–5390. https://doi.org/10.1109/tvt.2019.2908425

[124] Chuang-Wen You, Yaliang Chuang, Hung-Yeh Lin, Jui-Ting Tsai, Yi-Ching Huang, Chia-Hua Kuo, Ming-Chyi Huang, Shan Jean Wu, Frank Wencheng Liu, Jane Yung-Jen Hsu, and Hui-Ching Wu. 2019. SoberComm: Using Mobile Phones to Facilitate Inter-family Communication with Alcohol-dependent Patients. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (Sept. 2019), 1–31. https://doi.org/10.1145/3351277

[125] Chuang-Wen You, Ya-Fang Lin, Yaliang Chuang, Ya-Han Lee, Pei-Yi Hsu, Shih-Yao Lin, Chih-Chun Chang, Yi-Ju Chung, Yi-Ling Chen, Ming-Chyi Huang, Ping-Hsuan Shen, Hsin-Tung Tseng, and Hao-Chuan Wang. 2018. SoberMotion: Leveraging the Force of Probation Officers to Reduce the Risk of DUI Recidivism. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (Sept. 2018), 1–34. https://doi.org/10.1145/3264956

[126] Chuang-Wen You, Lu-Hua Shih, Hung-Yeh Lin, Yaliang Chuang, Yi-Chao Chen, Yi-Ling Chen, and Ming-Chyi Huang. 2019. Enabling Personal Alcohol Tracking using Transdermal Sensing Wristbands: Benefits and Challenges. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '19)*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/3338286.3344384

[127] Feng You, Xiaolong Li, Yunbo Gong, Hailwei Wang, and Hongyi Li. 2019. A Real-time Driving Drowsiness Detection Algorithm With Individual Differences Consideration. *IEEE Access* 7 (2019), 179396–179408. https://doi.org/10.1109/access.2019.2958667

[128] Ali Shahidi Zandi, Azhar Quddus, Laura Prest, and Felix J. E. Comeau. 2019. Non-Intrusive Detection of Drowsy Driving Based on Eye Tracking Data. *Transportation Research Record: Journal of the Transportation Research Board* 2673, 6 (May 2019), 247–257. https://doi.org/10.1177/0361198119847985

[129] Raimondas Zemblys, Christer Ahlström, Katja Kircher, and Svitlana Finér. 2024. Practical aspects of measuring camera-based indicators of alcohol intoxication in manual and automated driving. *IET Intelligent Transport Systems* 18, 8 (June 2024), 1408–1427. https://doi.org/10.1049/itr2.12520

A Participant details

Figure 13 illustrates the study flow diagram that outlines the recruitment process of our study, screening procedures, and subsequent group assignment.

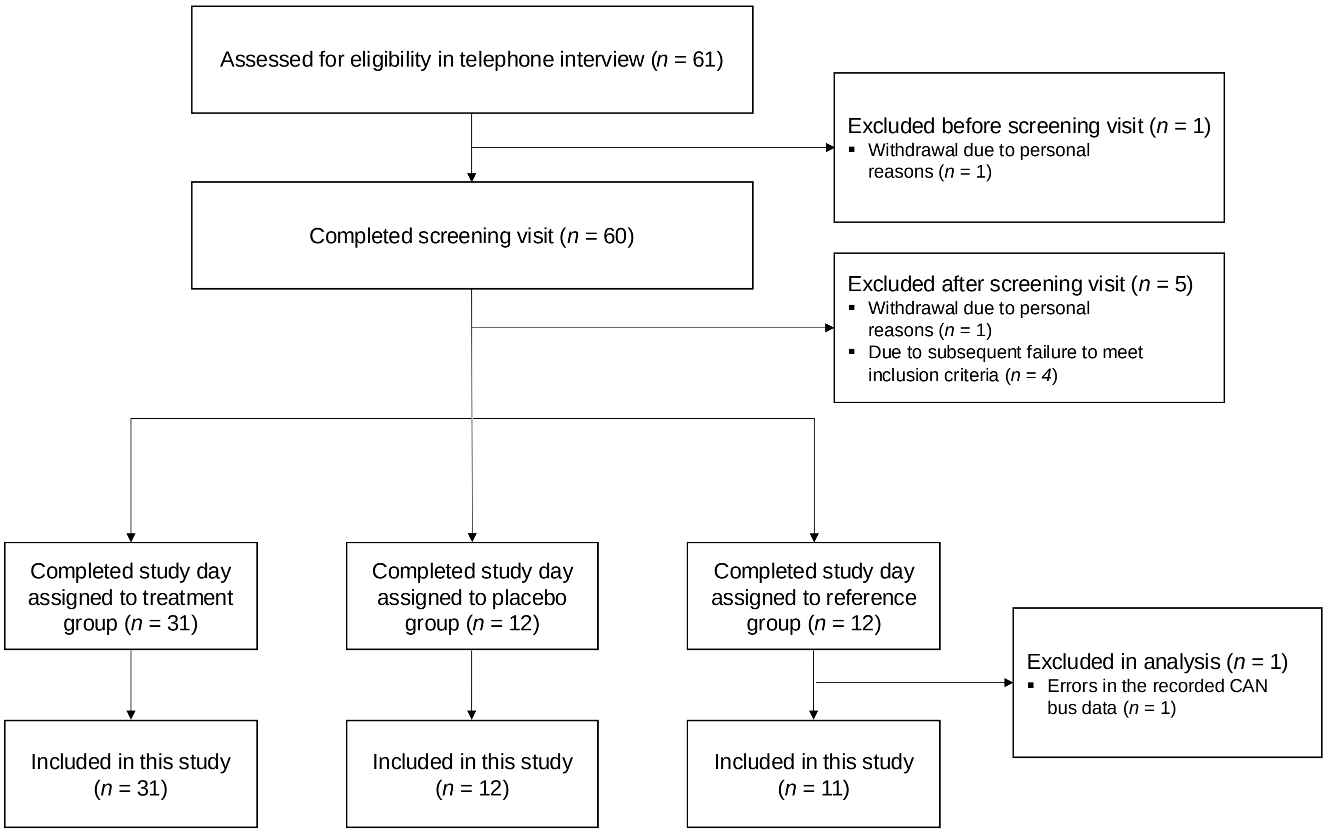

Figure 13: The figure displays our study flow diagram.

B Feature Generation

Figure 14 and 15 present comprehensive overviews of all features generated from DMC and CAN bus data, respectively.

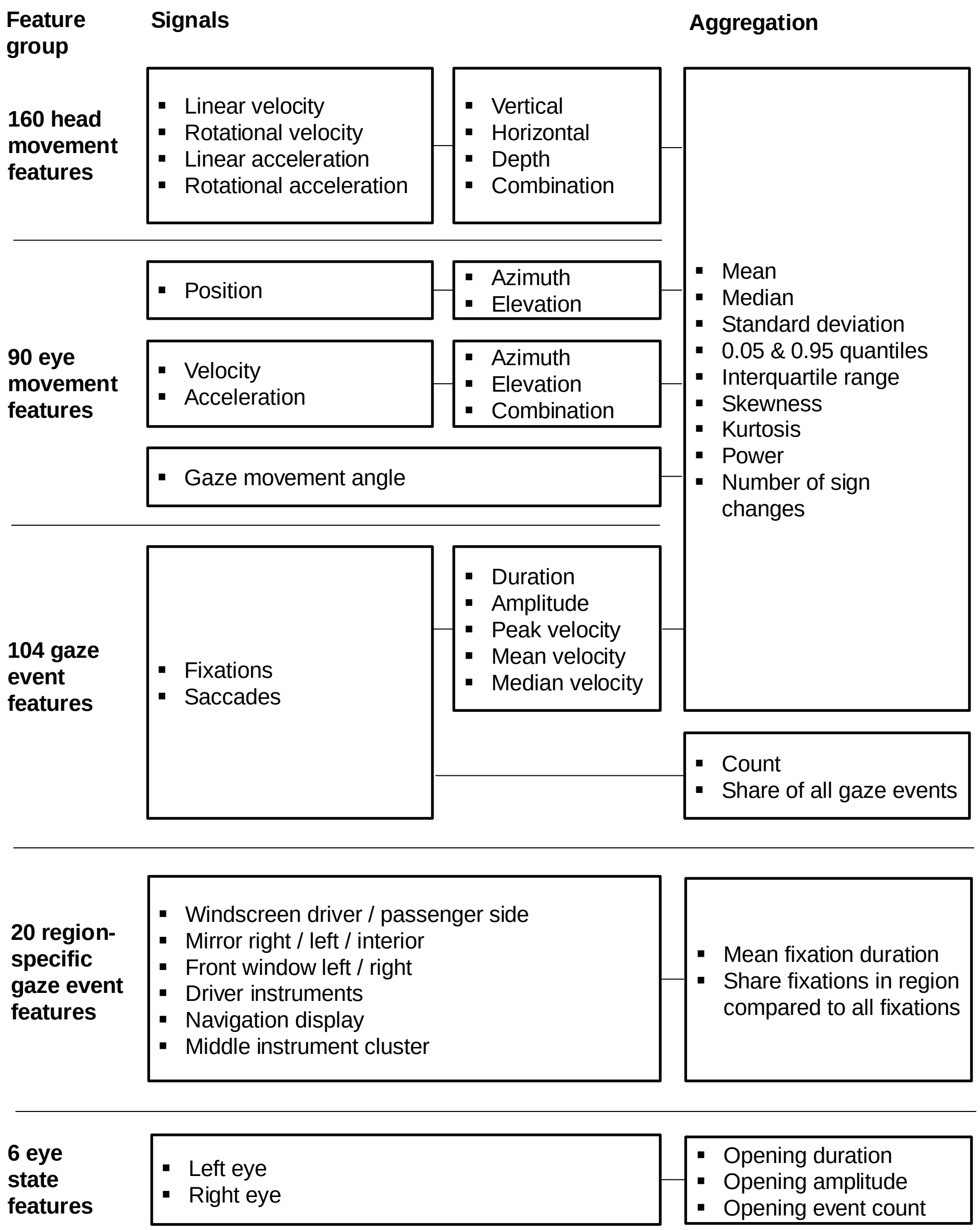

Figure 14: The figure displays an overview of all features generated from driver monitoring camera (DMC) data. We have five distinct feature groups, each based on unique base signals to which aggregation functions were subsequently applied.

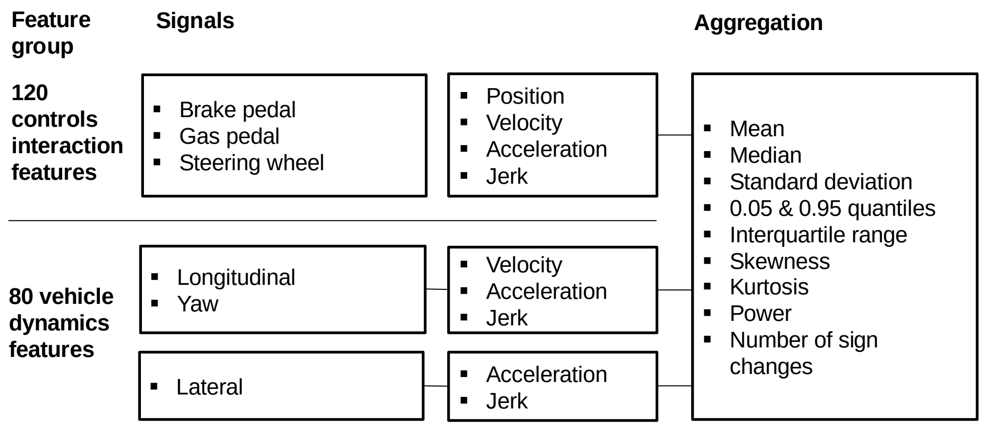

Figure 15: The figure displays an overview of all features generated from controller area network (CAN) sensor data. We have two distinct feature groups, each based on unique base signals to which aggregation functions were subsequently applied.