

PAPER

AI and Mobile Technologies for Driver Fatigue Detection: Sex Differences Revealed by Eye-Tracking Metrics

Julia Bend¹(✉), Markus
Gödker², Thomas Franke²

¹Åbo Akademi University,
Turku, Finland

²University of Lübeck,
Lübeck, Germany

julia.bend@abo.fi

ABSTRACT

The present study examines sex differences in fatigue and visual attention during simulated driving. Using a driving simulator, we evaluated the effects of task duration and sex differences on fatigue-related eye-tracking metrics, including blink duration, fixation rate, and blink rate. Results indicate that fixation rate was significantly influenced by task duration but remained consistent across sexes, whereas blink duration exhibited marginal sex differences and a significant interaction with task duration, with females demonstrating longer blink durations compared to males. These findings suggest that males and females adopt distinct patterns in managing fatigue over prolonged tasks, with implications for visual attention and fatigue. This study advances the understanding of sex-specific fatigue responses in dynamic tasks and underscores the potential of eye-tracking technologies for improving road safety.

KEYWORDS

driving, fatigue, visual attention, sex differences, eye tracking

1 INTRODUCTION

Fatigue is a critical factor impacting a wide range of activities, including driving, where its consequences can be severe [1]. In particular, driver drowsiness can lead to delayed reaction times, reduced attentiveness, and impaired information processing, which are key contributors to traffic accidents [2]. Due to its significant role in motor vehicle safety, driver drowsiness demands focused attention; in fact, impaired driving resulting from drowsiness and distraction contributes to nearly 40% of fatal crashes [3, 4]. Consequently, as understanding the psychophysiological effects of both internal and external demands on fatigue grows in importance, recent studies emphasize the need to explore fatigue as a result of loads that exceed an individual's coping capacity and tolerance levels [1, 5].

Furthermore, De La Vega et al. [1] identified significant differences in fatigue levels between male and female drivers in driving tasks. Their study revealed that females exhibited higher levels across all fatigue-related parameters evaluated,

Bend, J., Gödker, M., Franke, T. (2025). AI and Mobile Technologies for Driver Fatigue Detection: Sex Differences Revealed by Eye-Tracking Metrics. *International Journal of Interactive Mobile Technologies (IJIM)*, 19(11), pp. 143–158. <https://doi.org/10.3991/ijim.v19i11.53737>

Article submitted 2024-12-10. Revision uploaded 2025-03-03. Final acceptance 2025-03-13.

© 2025 by the authors of this article. Published under CC-BY.

including hours awake, heart rate, heart rate variability, Stroop Color-Word test scores, Samn-Perelli Scale ratings, Epworth questionnaire results, and overall stress levels. The authors attributed these differences to variations in stress response and regulation. However, they also noted that these disparities might be affected by additional factors beyond sex differences, such as activity levels, social background, education, and other elements that impact the human stress response [1]. Another study by Zeng et al. [6] also found sex differences in drivers' fatigue levels using heart rate variability. The authors suggested considering sex differences when developing driver fatigue detection systems that utilize heart rate variability features.

To understand cognitive and physiological states, various validated metrics are now available. For instance, electroencephalography and event-related potentials, cerebral blood flow imaging, blood pressure, electromyography, thermal imaging, heart rate variability, and pupillometry have proven effective in evaluating drivers' cognitive states [1, 7].

In driving, eyes play an essential role in performing visual tasks, such as road scanning, communication with other drivers, and monitoring vehicle indicators. As a result, eye-related parameters are frequently used to assess drivers' mental workload, fatigue, and visual attention [8, 9, 10]. Moreover, technological advances in camera resolution and computational power have made remote eye-tracking methods increasingly popular as non-intrusive and reliable tools for evaluating driver workload, leading to the development of innovative measurement techniques that can assess driver fatigue in real-time [9].

1.1 Blink metrics for fatigue assessment

Blinking, an involuntary reflex involving the rapid closure and reopening of the eyelids, serves essential functions such as spreading tears across the eye surface for lubrication and clearing debris. A blink, described as "a temporary closure of both eyes involving upper and lower eyelid movements," typically occurs around 12 times per minute in adults, with each blink lasting approximately one-third of a second [11, 12, 13].

Common parameters used in blink analysis are blink rate, blink duration, blink latency and the percentage of eye closing time over a certain period (PERCLOS) [9, 14, 15, 16]. Blink rate, defined as the number of blinks per minute or second, requires only the detection of blink occurrences, not the measurement of blink amplitude, duration, or other detailed characteristics. This simplicity might make blink rate one of the more accessible blink metrics to record. However, significant variability exists in blink rates both between individuals and over time within the same individual [17]. For instance, Zaman and Doughty [18] observed substantial fluctuations in blink rate over a five-minute recording session, suggesting that blink measurements may depend on the timing of data collection during an experiment. Research has frequently linked blink rate with mental workload and fatigue, underscoring its importance in cognitive assessment [9].

Additionally, Tsai et al. [19] found that drivers performing simultaneous auditory tasks exhibited higher blink rates compared to driving alone. Similarly, Van Orden et al. [20] reported that blink rate increased in response to tracking errors in human factors tasks. In a study by Li et al. [21] on operators' productivity and safety performance, fatigue was shown to increase blink rate while reducing pupil diameter and visual attention range. However, the relationship between blink rate and cognitive or visual load varies across studies. Blinks tend to be infrequent when visual demands are high, as they are associated with maintained concentration.

Therefore, the discrepancies in findings regarding blink rate reflect the nature of the task. For example, in piloting studies [22], the number of blinks decreased when pilots were required to process more visual information, regardless of the air traffic control task's demand level. However, when the task's requirements exceed cognitive resources, concentration may diminish, leading to an increase in blink rate as a result of higher mental workload [17].

PERCLOS has been widely associated with assessing attentional performance in tasks such as driving. Recent studies [15, 16] highlight its ability to differentiate between fatigue and alert states, where a PERCLOS value exceeding 15% indicates driver fatigue. The alignment of PERCLOS trends with real fatigue states, validated through heart rate measurements and subjective evaluations, underscores its reliability in monitoring driver attentiveness and identifying fatigue-induced performance impairments.

Blink rate and duration are also associated with performance in attentional tasks, such as driving or piloting, and show correlations with subjective sleepiness scales and EEG findings [23, 24]. During visually demanding tasks, such as reading, blink suppression or reduction may occur. Conversely, increased blink duration is often linked to higher levels of sleepiness or fatigue, making it a potentially more reliable alertness indicator than blink rate alone [24].

Further evidence of the link between blink duration and fatigue is provided by Mulhall et al. [25], who demonstrated that pre-drive blink duration could predict behavioral microsleeps and lane departures in shift workers. Their study found that longer blink durations were consistent predictors of sleep-related driving impairments, including self-reported and hazardous driving events. Similarly, Summala et al. [26] examined blink frequency and duration during overnight driving and found that blink rate increased with time-on-task, indicating rising fatigue, while blink suppression occurred when approaching an oncoming heavy vehicle, reflecting an attentive response to a potential hazard. This study supports previous findings that increased blink frequency with prolonged task duration reflects an effort to maintain attention when fatigued. The increasing need for objective and automated detection systems to monitor phenomena such as fatigue, drowsiness, and emotional states has led researchers to recognize that human blinking characteristics can be valuable indicators of cognitive state fluctuations [20, 27].

1.2 Fixation measures in fatigue detection

A "fixation" generally refers to a period when the gaze is directed at a small, specific area on the screen for a certain duration. Various techniques can be used to identify fixations from eye-tracking data, with a common method being the application of a gaze velocity threshold to classify fixations. However, fixation detection in head-mounted eye-tracking studies is significantly more challenging. In this study, we adopted a functional approach, considering any eye movement that stabilized a visual target on the retina as part of a fixation [28]. With this approach, microsaccades are included within fixations, meaning that fixations were not interrupted by these small eye movements [29].

Fixation rate is a measure calculated by dividing the total number of fixations by the trial duration in seconds, providing a rate in units per second commonly used in eye-tracking research. Previous research has observed that fixation rate tends to decrease as time on task increases [17, 30]. Additionally, Nakayama et al. [31] found a negative correlation between fixation rate and task difficulty, suggesting that the more challenging a task, the lower the fixation rate. Van Orden et al. [32] further noted that fixation rate varies with tracking performance across participants and demonstrated

its predictive value for target density in surveillance tasks, indicating its potential as a mental workload indicator [17]. Recent studies in driving, such as those by Weifeng et al. [33] and Sun et al. [34], have confirmed the validity of eye-tracking signals, particularly pupil size and fixation duration, as reliable measures for detecting eye fatigue.

In summary, there are various ways to assess a driver's visual attention and fatigue, but no single method is universally applicable. Driving is a complex task, and each measurement method has its specific limitations. While the above mentioned studies provide insights into the relationship between eye-tracking metrics and visual attention and fatigue, specific research focusing on sex differences in these parameters during driving tasks remains limited. Further investigation is needed to comprehensively understand how sex differences may influence eye-tracking measures and mental tiredness in driving contexts. The present study examines the following research question: "How do sex differences and driving task duration impact fatigue indicators such as fixations and blinks, and what are the potential implications for driver safety?" Understanding this interaction can reveal if females or males are more susceptible to fatigue over time in driving, which might influence attention loss or response times. Specific intervals or conditions where fatigue changes might predict reduced driving safety, helping tailor recommendations on optimal break times or alertness interventions based on both sex and task duration.

2 METHOD

2.1 Participants

A total of 105 participants completed the experiment. Recruitment was conducted through the university mailing list, forums on the university's learning management system, and social media platforms. Eligibility criteria required participants to be at least 18 years old, hold a valid driving license, and be fluent in German. Participation was voluntary, and informed consent was mandatory. Participants received compensation of €18.62 (aligned with the German minimum wage) or course credit for psychology and media informatics students at the University of Lübeck, but no additional performance-related reward. We decided to only compensate overall participation to avoid interindividual differences in motivation (and therefore in effort) due to the monetary effect. The study received ethics approval from the Ethics Committee of the University of Lübeck (tracking numbers 2023-680 and 2023-680_1).

Of all participants, four aborted the experiment due to simulation sickness, and eight had significant irregular procedures (e.g., major display failures, in compliant behavior), resulting in a sample of $N = 93$ participants. Of this sample, five participants had corrupted driving data sets due to temporary technical problems in the driving data logging. Due to technical issues with the eye-tracking recording setup, we could only include eye-tracking data from our sample after the introduction of a new eye-tracking apparatus. All analyzed data were collected using the new Pupil Labs' Pupil Neon eye-tracking glasses, ensuring consistency in the recording system and procedure for all participants included in the analysis. Hence, from the sample of $N = 88$ participants, only 53 participants used the functioning eye-tracking setup. Of these, three datasets were unusable because the participants wore incompatible glasses, and the final sample consisted of $N = 50$ participants.

All participants' ages ranged from 18 to 38 years ($M = 23.0$, $SD = 4.0$). In response to the question, "Which gender do you most closely identify with?" 36 participants identified as female (72%), 14 as male (28%), while no participants identified as diverse (0%), and none chose not to disclose their gender (0%). The mean total driving experience

with passenger vehicles (regardless of their type of drivetrain) was $M = 45876$ km ($SD = 144593$ km, $N = 48$; due to a technical error in the survey, we did not collect this information from every participant). 11 participants (22% of the sample) had already driven an EV at least 50 km with a total driving experience of $M = 801$ km ($SD = 1056$ km). We assessed the Affinity for Technology Interaction (ATI) scale [35] to characterize the sample regarding user diversity in human-technology interaction. The sample had a mean ATI score of $M = 3.92$ ($SD = 1.15$, possible range from 1 to 6), being higher than a German population quota sample (3.61, as described in Franke et al. [35]).

2.2 Apparatus

The study was conducted in the EcoSimLab electric vehicle driving simulator at the Institute of Multimedia and Interactive Systems, University of Lübeck. The simulated vehicle was a Renault Zoe EV (Electric Vehicle), and the driving environment used was the EcoDrivingTestPark [36]. The EcoDrivingTestPark is a driving environment with seven driving scenarios. The scenarios are designed to induce various energy-relevant maneuvers (see Figure 1 as an example), varied in situation complexity, and took on average $M = 77.5$ seconds ($SD = 10.0$, $Min = 55.8$, $Max = 123.6$). The hardware setup featured three 55-inch, 120 Hz monitors arranged to provide a 180-degree field of view. A single car seat was equipped with a Fanatec simulation rig, including a seat, wheelbase, and pedals. A compact car steering wheel was installed to replicate the appearance of a Renault Zoe. The steering wheel and pedals were enhanced with force feedback capabilities. The simulation software employed was BeamNG.tech, a specialized version of the BeamNG.drive simulation platform [37]. Energy consumption was modeled using an advanced energy model developed by Heidinger et al. [38] and further refined by Gödker et al. [36], accounting for vehicle and environmental physics within the simulation.

2.3 Procedure

In this repeated-measures driving simulator experiment, participants were randomly assigned to one of three display conditions: (1) optimal speed display (OSD), (2) instantaneous consumption trace display (ICD), and (3) no display/control (NOD). Each participant first completed both complexity blocks without any display (baseline phase) and then completed both blocks with the assigned display based on their group (experimental phase).

At the start of the experiment, participants were briefed on the procedure, provided informed consent, and completed an initial questionnaire. They were then given detailed information about the study through an instructional video. Participants familiarized themselves with the driving simulator by completing a tutorial track, lasting approximately four minutes, which included various speed limits, curves, inclines, and declines. After this, a second instructional video explained the tasks and rules of the EcoDrivingTestPark. Participants were instructed to drive as efficiently as possible while adhering to German traffic regulations. They were tasked with following traffic signs to navigate to the fictional city of “Simnitz” and avoid wrong turns. Each sector had a time limit, calculated based on the time it would take a driver maintaining 90% of the speed limit for that sector. This allowed participants to gauge their progress without knowing the exact time limit. Additionally, they were required to respond to online prompts when passing through teleportation tunnels.

Participants then completed a test trip (not part of the present analyses) on two sectors to become familiar with the EcoDrivingTestPark and their tasks. These test

sectors were not part of the core sectors for the experiment. Following the test trip, participants had to drive both complexity blocks (in randomized order and with a randomized order of sectors), hence, each sector once (baseline phases). After the baseline phase, participants watched a third instructional video explaining their assigned eco-driving display. The control group did not receive any additional instruction.

Participants then completed another test trip to practice using the display (not part of the analyses), followed by both complexity blocks (experimental phase, again driving each sector once).

Between the blocks, participants completed several questionnaires (not part of this study). After all trips were completed, they answered a final questionnaire on personal characteristics and demographics. Participants were then thanked and dismissed. The entire experimental procedure lasted approximately 90 minutes per participant.

Materials, measures, and eye-tracking apparatus. This experiment used Pupil Labs’ Pupil Neon eye-tracking glasses to record eye movements at a sampling rate of 200 Hz. The glasses featured an egocentric scene camera with a resolution of 1080 × 1088 pixels and a frame rate of 30 Hz. Two near-eye infrared (IR) cameras recorded each eye at a 200 Hz frame rate. Participants were allowed to move their heads freely during the experiment, and a head strap was used to prevent the glasses from slipping. The primary dependent variables in this study were eye blink duration, fixation rate, and blink rate.

Blink and fixation detection and data processing. To confidently identify blinks, we included only timepoints where both eye tracks were missing. Data were limited to samples collected while participants were actively performing tasks, excluding screens for instructions. The dataset was further filtered to retain only samples with both eyes undetected.



Fig. 1. Driving environment with seven driving scenarios

For blink detection, lower and upper limits on undetected pupil duration were applied. The lower threshold for blink detection was set at 100 ms, based on the

Pupil Labs blink detection algorithm, which aligns with physiological blink limits reported in the literature. This threshold is widely accepted as it minimizes the risk of misclassifying shorter interruptions as blinks. Data losses exceeding 1000 ms were excluded as they were attributed to microsleeps or other non-task-related interruptions, ensuring the integrity of blink-related measures.

For fixation detection, the basic velocity-based I-VT algorithm classifies each gaze sample as either a fixation or non-fixation, depending on whether gaze speed is below or above a set threshold [39, 40]. Although effective under static viewing conditions with minimal head motion, this method faces challenges in head-mounted eye-tracking studies where head movement is present. To address this, we used Pupil Labs' adaptive velocity threshold algorithm [29]. For this study, the minimum velocity threshold was set at $v_0 = 1200$ px/s with a gain factor of $g = 0.8$, and a minimum fixation duration of $d_{min} = 70$ ms. As a post-processing step, fixations separated by gaps smaller than 2° in amplitude and shorter than 50 ms (microsaccades) were merged into fixations, ensuring they were included as uninterrupted events [29]. No maximum fixation duration limit was applied.

2.4 Justification for thresholds

The thresholds selected for blink and fixation metrics were based on a review of relevant literature and the limitations of the recording equipment. For blink duration, the 100 ms threshold reflects natural physiological limits and aligns with settings used in similar studies. Many researchers adopt thresholds in the range of 80–100 ms as they align better with typical blink rates, though these settings may not suit all studies. For example, Hollander and Huette [12] used 160 ms; Frank et al. [41] and Huette et al. [42] applied a 100 ms threshold; and Van Orden et al. [32] used 83.3 ms. The Pupil Labs blink detection algorithm uses a default lower threshold of 100 ms, which we adopted in this study. This threshold effectively captures most spontaneous blinks while excluding shorter interruptions, such as noise or minor occlusions, enhancing the reliability and validity of our findings.

3 RESULTS

For all trials, data were analyzed using mixed-effects ANOVA with a between-subject factor of sex (male/female) and a within-subjects factor of trials (4 levels). Figures 2, 3, and 4 plot average blink duration, fixation rate, and blink rate, with males and females represented by blue and red bars, respectively. Table 1 provides descriptive statistics for males and females across trials.

For *fixation rate*, ANOVA showed a significant main effect of *trial* $F(3,144) = 3.671$, $p = .014$, $p < 0.1$, $\eta^2 = 0.071$. Neither the main effect of *sex*, $F(1,48) = 0.080$, $p = .778$, $\eta^2 = 0.002$, nor the *sex*trial* interaction, $F(3,144) = 0.209$, $p = .890$, $\eta^2 = 0.004$, were significant. The significant effect of Trial suggests systematic changes in fixation rates over time, likely reflecting task progression. This aligns with prior findings that fixation behavior is influenced by task demands.

For *blink rate* as a dependent variable, the main effect of *trial* $F(3,144) = 2.583$, $p = .056$, $p < 0.1$, $\eta^2 = 0.051$ was marginally significant, whereas both *sex* $F(1,48) = 0.013$, $p = .909$, $\eta^2 = 0.000$ and the *sex*trial* interaction $F(3,144) = 1.573$, $p = .199$, $\eta^2 = 0.032$ were not. The marginal trial effect may reflect changes in participants' fatigue or adaptation to the task. Although sex differences were not observed, the role of blink rate as an indicator of fatigue requires further exploration.

Finally, for *average blink duration* as a dependent variable, mixed-effects ANOVA showed a marginally significant effect of the *sex*trial* interaction, $F(3,144) = 2.628$, $p = .053$, $p < 0.1$, $\eta^2 = 0.052$, and an effect that approached but did not reach significance of *sex* $F(1,48) = 3.296$, $p = .076$, $p < 0.1$, $\eta^2 = 0.064$. The effect of *trial* $F(3,144) = 1.443$, $p = .233$, $\eta^2 = 0.029$ was not significant. Although these findings do not meet strict significance thresholds, they suggest potential sex-based differences in fatigue-related behavior. Females showed longer blink durations, which needs further exploration with a larger group of participants.

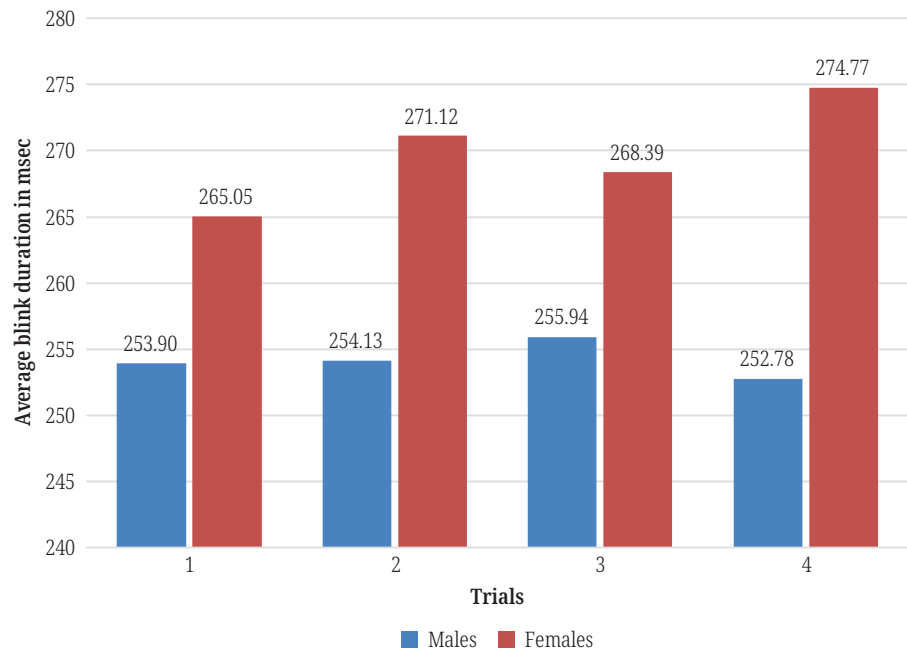


Fig. 2. Average blink duration

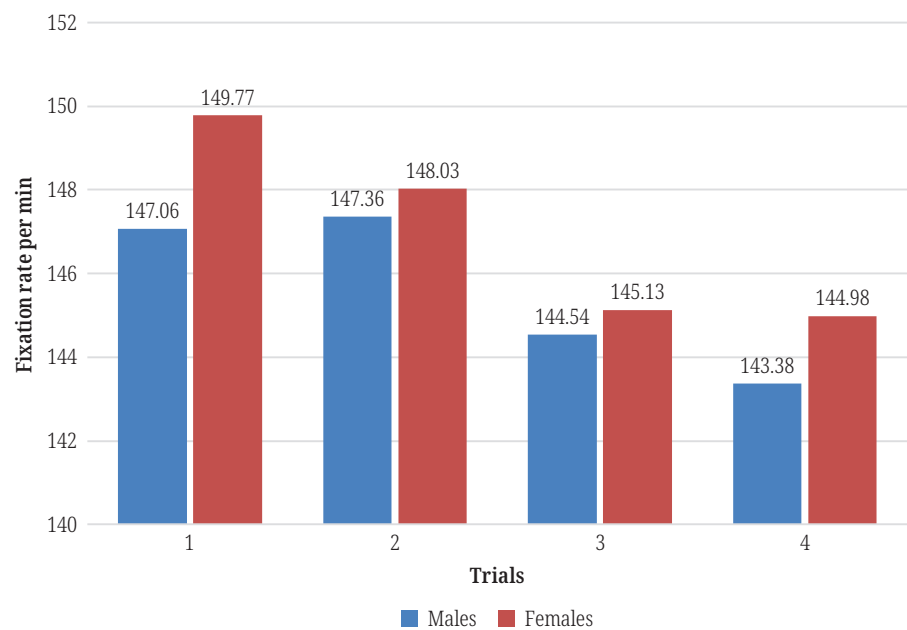


Fig. 3. Fixation rate per minute

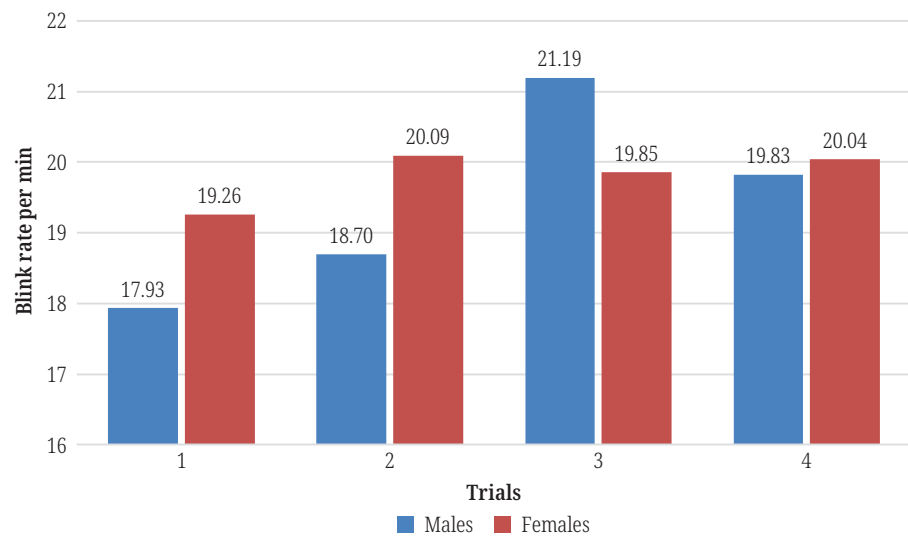


Fig. 4. Blink rate per minute

The term “marginally significant” is used to describe results where p-values slightly differ from the .05 threshold. Although these findings do not meet conventional significance criteria, they warrant discussion due to their potential practical implications. For example, the marginal interaction between blink duration and trial could reflect subtle differences in how males and females respond to fatigue over time.

Table 1. Descriptive statistics for males and females

Sex	Eye Tracking Measurement	Parameter	Trial 1	Trial 2	Trial 3	Trial 4
Males		Valid N	14	14	14	14
	Fixation rate per minute	Mean	147.06 (14.70)	147.36 (19.40)	144.54 (16.97)	143.38 (17.45)
	Blink rate per minute	Mean	17.93 (9.10)	18.7 (9.89)	21.19 (13.85)	19.83 (12.23)
	Average blink duration	Mean	253.9 (29.01)	254.13 (28.57)	255.94 (22.84)	252.78 (31.20)
Females		Valid N	36	36	36	36
	Fixation rate per minute	Mean	149.77 (16.78)	148.03 (16.60)	145.13 (16.35)	144.98 (16.40)
	Blink rate per minute	Mean	19.26 (10.99)	20.09 (11.74)	19.85 (11.70)	20.04 (10.64)
	Average blink duration	Mean	265.05 (26.77)	271.12 (26.57)	268.39 (31.00)	274.77 (30.55)

Table 1 presents the mean (std. dev.) for each eye-tracking metric across four trials, separated by sex (males and females).

4 DISCUSSION

Fatigue from extended mental activity and other contributing factors is a key cause of traffic accidents and a significant public health issue [1, 2]. Addressing

fatigue holds relevance across various scenarios, particularly in driving, where tiredness can have severe consequences [1]. With increasing automation in vehicles capable of managing speed and lane-keeping, diagnosing driver impairment has become more challenging because automation limits the availability of vehicle-based signals (such as steering input or braking behavior) traditionally used to assess driver performance. This necessitates the use of camera-based driver monitoring systems (DMS) [4]. Moreover, when the perceived effort or fatigue outweighs the expected reward, individuals are less motivated to sustain task performance [1]. Maintaining accurate decision-making in dynamic environments depends heavily on sustained attention, which can be compromised by mental fatigue arising from prolonged cognitive efforts [43]. Consequently, precise assessments of fatigue, particularly in its mental dimension, are crucial for improving road safety, especially as automation increasingly places drivers in supervisory roles. The advancement of camera technologies and computational capabilities has facilitated the use of remote eye measurements as unobtrusive tools to assess driver workload effectively [9]. The present study examines how sex differences and driving task duration impact fatigue indicators such as fixations and blinks, with potential implications for driver safety. Specifically, we investigated whether males and females differ in baseline fatigue indicators, how prolonged driving influences fatigue, and whether sex and task duration interact to shape fatigue-related changes.

While various approaches exist to evaluate mental workload, no single method is universally effective, given the complexity and multidimensional nature of driving tasks, each with inherent limitations. Research has documented sex-based differences in driving-related fatigue using heart rate variability metrics [1]; however, these differences have not yet been examined through eye-tracking methodologies. This may be due to the fact that physical fatigue measures have undergone more comprehensive development in recent years. For example, in sports and physical activity contexts, heart rate variability metrics have been shown to be effective for monitoring fatigue across varying training intensities [1]. In our study, we employed eye-tracking to assess drivers' fixation patterns and blink behaviors during a driving task as indicators of fatigue levels. Marginally significant effects of sex and the sex*task duration interaction on *blink duration* were found. This suggests a potential tendency for sex differences in blink duration, with females appearing to blink for longer periods than males. Although this effect is not highly significant, it may indicate that males and females exhibit different blink behaviors. The interaction suggests that the effect of task duration on blink duration was different for males and females. The significance indicates that as the task duration increased, males and females diverged in how their blink duration changed over time, possibly due to differences in how they handled fatigue. Task duration significantly impacted *fixation rates* and had a marginal effect on *blink rates*, consistent with prior studies associating these metrics with fatigue over time [9, 15, 16, 17, 18, 20, 21].

Implications. Our findings contribute to the theoretical understanding of fatigue regulation by highlighting sex differences in blink duration, suggesting that males and females may experience and respond to extended driving tasks differently. This underscores the need for tailored fatigue mitigation strategies. Previous research has emphasized the role of sex-specific factors in fatigue detection, with Zeng et al. [6] recommending their integration into heart rate variability-based monitoring systems for improved accuracy. Similarly, Lin and Chien [44] demonstrated that sex influences operational performance and user satisfaction with car navigation systems, advocating for interface designs that minimize distractions. The present study extends this work by providing behavioral insights through eye-tracking metrics, particularly in the use of blink duration as a potential sex-sensitive fatigue marker.

From an applied perspective, sex differences in fatigue regulation highlight the need for personalized, artificial intelligence (AI)-driven fatigue mitigation strategies. Eye-tracking-based systems, as noted by [45], currently achieve up to 95% accuracy, outperforming respiration-based (90%) and facial feature-based (74%) systems. Many modern DMS, such as those by smart eye AB and driver state sensors (DSS) by Seeing Machines, rely on AI and computer vision to analyze facial and eye-tracking metrics [2, 26]. To improve fatigue detection, comprehensive systems integrating AI-based models with multiple data sources are essential.

According to [46], modern fatigue detection systems leverage multi-sensor, smartphone-based, and cloud-based architectures, analyzing eye-tracking data, heart rate variability (HRV), and vehicle movement patterns for real-time fatigue prediction. Additionally, machine learning frameworks, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have been widely adopted to classify fatigue levels based on visual cues such as PERCLOS, blink frequency, and head movements [45]. These models enable low-latency fatigue detection through edge computing and 5G networks, improving response times for real-time applications. Yang et al. (2024) [47] further highlight how AI-powered DMS leverage deep learning-based gaze tracking, head pose estimation, and blinking analysis to assess driver attention and drowsiness. These AI-driven solutions integrate adaptive fatigue warnings and intervention strategies, enhancing real-time safety mechanisms. While wearable sensors and EEG-based systems offer high accuracy, smartphone-based fatigue monitoring presents a cost-effective alternative for large-scale deployment.

However, various approaches to fatigue monitoring in driving and occupational settings highlight the importance of accurate and timely detection [48]. Wearable eye-tracking technology has shown promising performance in mental fatigue detection, achieving high classification accuracy using widely adopted machine learning algorithms. For example, [21] revealed that combining pupil and blink features resulted in the highest classification accuracy, suggesting that even low-cost devices such as webcams can be utilized for automatic fatigue monitoring. This makes eye-tracking-based methods accessible for budget-constrained projects. However, it is important to recognize the limitations of behavior-based methods, as they primarily detect fatigue after its onset. This delayed detection may limit their ability to prevent exposure to fatigue-related risks. Additionally, computer vision-based systems remain sensitive to environmental factors, such as lighting variability, which can impact accuracy [10].

4.1 Future directions

To improve fatigue risk management, detecting early transitions into fatigue is critical, rather than focusing solely on post-fatigue detection. AI-driven models offer a significant advantage by identifying fatigue onset before it affects cognitive performance and safety [21]. One such AI-based method, the Toeplitz Inverse Covariance-Based Clustering (TICC) algorithm, enables real-time classification of fatigue levels based on both objective (eye-tracking metrics) and subjective (self-reported) data. Unlike traditional threshold-based detection, adaptive AI models continuously learn from data, detecting subtle transitions between fatigue states [21].

However, while AI-driven fatigue detection systems enhance driver safety, they also raise critical ethical and privacy concerns, particularly regarding data collection, usage, transparency, accountability, and regulatory oversight. Wani et al. [49] highlight the need for privacy-preserving AI techniques and ethical frameworks to balance user rights and safety in autonomous driving applications. Additionally,

eye-tracking data is susceptible to privacy risks, including re-identification and inference attacks. Seyedi et al. [50] demonstrated that deep learning models trained on full-face video recordings may inadvertently expose sensitive biometric data, underscoring the urgent need for robust security measures and regulatory guidelines.

4.2 Limitations

The present study provides valuable insights into sex differences in driver fatigue; however, several limitations must be acknowledged to contextualize the findings and guide future research.

First, the final sample size of 50 participants, with a skewed gender distribution (72% female), limits the generalizability of the results. While the study highlights potential sex-based differences in fatigue indicators, the uneven sample distribution and relatively small size may reduce the statistical power to detect subtle differences. Future studies should aim to recruit a more balanced and larger sample to enhance the robustness and generalizability of findings, particularly regarding sex-specific differences.

Second, the term “marginally significant” is used in the manuscript to describe results with p-values close to 0.05. While these findings are acknowledged, their practical implications require deeper exploration. Future research should aim to incorporate larger datasets and additional statistical analyses to validate these trends.

Third, although the study adequately addressed ethical considerations, including participant consent and ethics committee approval, the potential influence of the compensation method warrants further discussion. Participants received compensation for their overall participation rather than performance to minimize variability in motivation. However, this approach may still have influenced participants' behavior (e.g., completing tasks despite fatigue). Future studies should carefully consider how compensation methods might impact participant effort and task engagement to ensure transparency and reproducibility.

Lastly, the exploratory nature of this study and the use of a driving simulator present inherent limitations. Simulators, while valuable for controlled experiments, may not fully capture real-world driving conditions, potentially impacting the external validity of the findings. Future research should consider supplementing simulator studies with real-world assessments to enhance ecological validity.

By addressing these limitations and incorporating additional methodologies, future studies can build upon these findings to advance the understanding of sex differences in fatigue and their implications for driver safety technologies.

5 CONCLUSIONS

This study highlights differences in fatigue levels between females and males during driving tasks. Understanding these variations and their impact on driving performance is crucial for the development of advanced monitoring systems aimed at detecting and preventing accidents. Future research should aim to expand on this work by conducting longer experiments, incorporating a wider variety of driving scenarios, and increasing the sample size to ensure more robust and generalizable findings. Combining eye-tracking data with heart rate variability measurements could further enhance our understanding of fatigue and drowsiness in both female and male drivers. Our findings suggest that prolonged driving may affect females

and males differently in terms of fatigue and drowsiness, offering valuable insights for improving driver safety.

6 ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG) under the name “AMORi” (project number 498999989), by the German Federal Ministry of Education and Research (BMBF) under the name “HI.Lübeck” (grant number 03IHS245B), and by the Åbo Akademi University Foundation and Gösta Branders research fund. The authors would like to thank Anssi Öörni, Egor Bend, Jan Heidinger, Lukas Bernhardt, Leonardt Wagner, Tobias Harms, Elise Banach, and Alexander Müller for their contributions to the success of this research.

7 REFERENCES

- [1] R. De La Vega, H. Anabalon, K. Tannion, H. Purto, and D. C. Jara, “Gender differences in professional drivers’ fatigue level measured with BALert mobile app: A psychophysiological, time efficient, accessible, and innovative approach to fatigue management,” *Front. Psychol.*, vol. 13, p. 953959, 2022. <https://doi.org/10.3389/fpsyg.2022.953959>
- [2] M. Abulhair *et al.*, “Mobile platform detect and alerts system for driver fatigue,” *Procedia Computer Science*, vol. 62, pp. 555–564, 2015. <https://doi.org/10.1016/j.procs.2015.08.531>
- [3] G. Zhang, K. K. Yau, X. Zhang, and Y. Li, “Traffic accidents involving fatigue driving and their extent of casualties,” *Accident Anal. Prevention*, vol. 87, pp. 34–42, 2016. <https://doi.org/10.1016/j.aap.2015.10.033>
- [4] C. Schwarz, J. Gaspar, T. Miller, and R. Yousefian, “The detection of drowsiness using a driver monitoring system,” *Traffic Injury Prevention*, vol. 20, no. sup1, pp. 157–161, 2019. <https://doi.org/10.1080/15389588.2019.1622005>
- [5] R. M. Enoka and J. Duchateau, “Translating fatigue to human performance,” *Med. Sci. Sports Exercise*, vol. 48, no. 11, pp. 2228–2238, 2016. <https://doi.org/10.1249/MSS.0000000000000929>
- [6] C. Zeng *et al.*, “Driver fatigue detection using heart rate variability features from 2-minute electrocardiogram signals while accounting for sex differences,” *Sensors*, vol. 24, no. 13, p. 4316, 2024. <https://doi.org/10.3390/s24134316>
- [7] M. Lohani, B. R. Payne, and D. L. Strayer, “A review of psychophysiological measures to assess cognitive states in real-world driving,” *Front. Hum. Neurosci.*, vol. 13, 2019. <https://doi.org/10.3389/fnhum.2019.00057>
- [8] K. A. Brookhuis and D. De Waard, “Monitoring drivers’ mental workload in driving simulators using physiological measures,” *Accident Analysis and Prevention*, vol. 42, no. 3, pp. 898–903, 2010. <https://doi.org/10.1016/j.aap.2009.06.001>
- [9] G. Marquart, C. Cabrall, and J. de Winter, “Review of eye-related measures of drivers’ mental workload,” *Procedia Manufacturing*, vol. 3, pp. 2854–2861, 2015. <https://doi.org/10.1016/j.promfg.2015.07.783>
- [10] N. R. Adão Martins, S. Annaheim, C. M. Spengler, and R. M. Rossi, “Fatigue monitoring through wearables: A state-of-the-art review,” *Front. Physiol.*, vol. 12, p. 790292, 2021. <https://doi.org/10.3389/fphys.2021.790292>
- [11] K. A. Kwon *et al.*, “High-speed camera characterization of voluntary eye blinking kinematics,” *J. R. Soc. Interface*, vol. 10, no. 85, p. 20130227, 2013. <https://doi.org/10.1098/rsif.2013.0227>
- [12] J. Hollander and S. Huette, “Extracting blinks from continuous eye-tracking data in a mind wandering paradigm,” *Consciousness and Cognition*, vol. 100, p. 103303, 2022. <https://doi.org/10.1016/j.concog.2022.103303>

- [13] I. Fatt and B. Weissman, *Physiology of the Eye. An Introduction to the Vegetative Functions* (2nd ed.). London: Butterworth-Heinemann, 2013.
- [14] T. Halverson, J. Estep, J. Christensen, and J. Monnin, "Classifying workload with eye movements in a complex task," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 56, no. 1, pp. 168–172, 2012. <https://doi.org/10.1177/1071181312561012>
- [15] M. Zhang, X. Zhai, G. Zhao, T. Chong, and Z. Wang, "An application of particle swarm algorithms to optimize hidden markov models for driver fatigue identification," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 25–30. <https://doi.org/10.1109/IVS.2018.8500357>
- [16] J. Xie, M. Xie, and W. Zhu, "Driver fatigue detection based on head gesture and PERCLOS," in *2012 IEEE International Conference on Wavelet Active Media Technology and Information Processing*, 2012, pp. 128–131. <https://doi.org/10.1109/ICWAMTIP.2012.6413456>
- [17] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van De Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford: Oxford University Press, 2011.
- [18] M. L. Zaman and M. J. Doughty, "Some methodological issues in the assessment of the spontaneous eyeblink frequency in man," *Ophthalmic and Physiological Optics*, vol. 17, no. 5, pp. 421–432, 1997.
- [19] Y. F. Tsai, E. Viirre, C. Strychacz, B. Chase, and T. P. Jung, "Task performance and eye activity: Predicting behavior relating to cognitive workload," *Aviation, Space, and Environmental Medicine*, vol. 78, no. 5 Suppl, pp. B176–B185, 2007.
- [20] K. F. Van Orden, T. P. Jung, and S. Makeig, "Combined eye activity measures accurately estimate changes in sustained visual task performance," *Biological Psychology*, vol. 52, no. 3, pp. 221–240, 2000. [https://doi.org/10.1016/S0301-0511\(99\)00043-5](https://doi.org/10.1016/S0301-0511(99)00043-5)
- [21] J. Li *et al.*, "Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology," *Autom. Constr.*, vol. 109, p. 103000, 2020. <https://doi.org/10.1016/j.autcon.2019.103000>
- [22] J. A. Veltman and A. W. K. Gaillard, "Physiological indices of workload in a simulated flight task," *Biological Psychology*, vol. 42, no. 3, pp. 323–342, 1996. [https://doi.org/10.1016/0301-0511\(95\)05165-1](https://doi.org/10.1016/0301-0511(95)05165-1)
- [23] J. B. Brookings, G. F. Wilson, and C. R. Swain, "Psychophysiological responses to changes in workload during simulated air traffic control," *Biol. Psychol.*, vol. 42, no. 3, pp. 361–377, 1996. [https://doi.org/10.1016/0301-0511\(95\)05167-8](https://doi.org/10.1016/0301-0511(95)05167-8)
- [24] Y. Wang, S. S. Toor, R. Gautam, and D. B. Henson, "Blinking frequency and duration during Perimetry and their relationship to test-retest threshold variability," *Investigative Ophthalmology & Visual Science*, vol. 52, no. 7, pp. 4546–4550, 2011. <https://doi.org/10.1167/iovs.10-6553>
- [25] M. D. Mulhall *et al.*, "A pre-drive ocular assessment predicts alertness and driving impairment: A naturalistic driving study in shift workers," *Accident Analysis & Prevention*, vol. 135, p. 105386, 2020. <https://doi.org/10.1016/j.aap.2019.105386>
- [26] H. Summala, H. Hakkanen, T. Mikkola, and J. Sinkkonen, "Task effects on fatigue symptoms in overnight driving," *Ergonomics*, vol. 42, no. 6, pp. 798–806, 2010. <https://doi.org/10.1080/001401399185298>
- [27] S. Chen and J. Epps, "Automatic classification of eye activity for cognitive load measurement with emotion interference," *Computer Methods and Programs in Biomedicine*, vol. 110, no. 2, pp. 111–124, 2013. <https://doi.org/10.1016/j.cmpb.2012.10.021>
- [28] R. S. Hessels, D. C. Niehorster, M. Nyström, R. Andersson, and I. T. C. Hooge, "Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers," *Royal Society Open Science*, vol. 5, no. 8, p. 180502, 2018. <https://doi.org/10.1098/rsos.180502>

- [29] M. Drews and K. Dierkes, "Strategies for enhancing automatic fixation detection in head-mounted eye tracking," *Behavior Research Methods*, vol. 56, pp. 6276–6298, 2024. <https://doi.org/10.3758/s13428-024-02360-0>
- [30] D. K. McGregor and J. A. Stern, "Time on task and blink effects on saccade duration," *Ergonomics*, vol. 39, no. 4, pp. 649–660, 1996. <https://doi.org/10.1080/00140139608964487>
- [31] M. Nakayama, K. Takahashi, and Y. Shimizu, "The act of task difficulty and eye movement frequency for the 'Oculomotor indices'," in *Proceedings of the 2002 Symposium on Eye-Tracking Research & Applications*, 2002, pp. 37–42. <https://doi.org/10.1145/507072.507080>
- [32] K. F. Van Orden, W. Limbert, S. Makeig, and T. P. Jung, "Eye activity correlates of workload during a visuospatial memory task," *Human Factors*, vol. 43, no. 1, pp. 111–121, 2001. <https://doi.org/10.1518/001872001775992570>
- [33] S. Weifeng, W. Yuqi, H. Bingliang, and W. Quan, "Exploring the connection between eye movement parameters and eye fatigue," in *Proceedings of the 2023 The 4th Asia Conference on Computers and Communications (ACCC 2023)*, 2023.
- [34] W. Sun, Y. Wang, B. Hu, and Q. Wang, "Exploration of eye fatigue detection features and algorithm based on eye-tracking signal," *Electronics*, vol. 13, no. 10, p. 1798, 2024. <https://doi.org/10.3390/electronics13101798>
- [35] T. Franke, C. Attig, and D. Wessel, "A personal resource for technology interaction: Development and validation of the Affinity for Technology Interaction (ATI) scale," *International Journal of Human-Computer Interaction*, vol. 35, no. 6, pp. 456–467, 2019. <https://doi.org/10.1080/10447318.2018.1456150>
- [36] M. Gödker, S. Schmees, L. Bernhardt, J. Heidinger, D. Görge, and T. Franke, "Driving simulation for energy efficiency studies: Analyzing electric vehicle eco-driving with EcoSimLab and the EcoDrivingTestPark," in *Proceedings of the 16th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '24)*, 2024, pp. 32–42. <https://doi.org/10.1145/3640792.3675706>
- [37] BeamNG GmbH., *BeamNG.tech* (Version 0.25.0.0) [Software]. BeamNG GmbH. <https://www.beamng.tech/research/>
- [38] J. Heidinger, L. Bernhardt, and T. Franke, "EcoSimLab—A low-cost driving simulator environment for examining human factors in vehicle energy efficiency," in *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 2023, pp. 66–70. <https://doi.org/10.1145/3581961.3609881>
- [39] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, 2000, pp. 71–78. <https://doi.org/10.1145/355017.355028>
- [40] R. Andersson, L. Larsson, K. Holmqvist, M. Stridh, and M. Nyström, "One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms," *Behavior Research Methods*, vol. 49, no. 2, pp. 616–637, 2017. <https://doi.org/10.3758/s13428-016-0738-9>
- [41] D. J. Frank, B. Nara, M. Zavagnin, D. R. Touron, and M. J. Kane, "Validating older adults' reports of less mind-wandering: An examination of eye movements and dispositional influences," *Psychology and Aging*, vol. 30, no. 2, pp. 266–278, 2015. <https://doi.org/10.1037/pag0000031>
- [42] S. Huette, A. Mathis, and A. Graesser, "Blink durations reflect mind wandering during reading," *Cog. Sci.*, 2016. <https://pdfs.semanticscholar.org/fba2/314dea3944723ddc-0348dc41a6e823dd5410.pdf>
- [43] S. M. Marcora, W. Staiano, and V. Manning, "Mental fatigue impairs physical performance in humans," *J. Appl. Physiol.*, vol. 106, no. 3, pp. 857–864, 2009. <https://doi.org/10.1152/jappphysiol.91324.2008>

- [44] P. C. Lin and L. W. Chien, "The effects of gender differences on operational performance and satisfaction with car navigation systems," *Int. J. Human-Computer Studies.*, vol. 68, no. 10, pp. 777–787, 2010. <https://doi.org/10.1016/j.ijhcs.2010.06.006>
- [45] F. Makhmudov, D. Turimov, M. Xamidov, F. Nazarov, and Y. I. Cho, "Real-time fatigue detection algorithms using machine learning for yawning and eye state," *Sensors*, vol. 24, p. 7810, 2024. <https://doi.org/10.3390/s24237810>
- [46] Q. Abbas and A. Alsheddy, "Driver fatigue detection systems using multi-sensors, smart-phone, and cloud-based computing platforms: A comparative analysis," *Sensors*, vol. 21, no. 1, p. 56, 2021. <https://doi.org/10.3390/s21010056>
- [47] G. Yang, C. Ridgeway, A. Miller, and A. Sarkar, "Comprehensive assessment of artificial intelligence tools for driver monitoring and analyzing safety critical events in vehicles," *Sensors*, vol. 24, no. 8, p. 2478, 2024. <https://doi.org/10.3390/s24082478>
- [48] S. A. El-Nabi, W. El-Shafai, E. S. M. El-Rabaie, K. F. Ramadan, F. E. Abd El-Samie, and S. Mohsen, "Machine learning and deep learning techniques for driver fatigue and drowsiness detection: A review," *Multimed. Tools Appl.*, vol. 83, pp. 9441–9477, 2024. <https://doi.org/10.1007/s11042-023-15054-0>
- [49] A. Wani, D. Kumari, and J. Singh, "Ethics in the driver's seat: Unravelling the ethical dilemmas of AI in autonomous driving," SAE Technical Paper, 2024.
- [50] S. Seyedi *et al.*, "An investigation of privacy preservation in deep learning-based eye-tracking," *BioMed. Eng. OnLine*, vol. 21, no. 67, 2022. <https://doi.org/10.1186/s12938-022-01035-1>

8 AUTHORS

Julia Bend is with the Åbo Akademi University, Turku, Finland (E-mail: julia.bend@abo.fi).

Markus Gödker is with the University of Lübeck, Lübeck, Germany.

Thomas Franke is with the University of Lübeck, Lübeck, Germany.