Transportation Research Interdisciplinary Perspectives 26 (2024) 101148

Contents lists available at ScienceDirect

Transportation Research Interdisciplinary Perspectives

journal homepage: www.sciencedirect.com/journal/transportation-research-interdisciplinary-perspectives

Classification of driver fatigue in conditionally automated driving using physiological signals and machine learning

Quentin Meteier [a,*], Reńee Favre [b], Sofia Viola [b], Marine Capallera [a], Leonardo Angelini [a,c], Elena Mugellini [a], Andreas Sonderegger [b,d]

[a] HumanTech Institute, University of Applied Sciences and Arts of Western Switzerland, HES-SO, Boulevard de Ṕerolles 80, Fribourg 1700, Switzerland
[b] Department of Psychology, University of Fribourg, Rue P.-A.-de-Faucigny 2, Fribourg 1700, Switzerland
[c] School of Management Fribourg, University of Applied Sciences and Arts of Western Switzerland, HES-SO, Chemin du Muśee 4, Fribourg 1700, Switzerland
[d] Bern University of Applied Sciences, Business School, Institute for New Work, Brückenstrasse 73, Bern 3005, Switzerland

A R T I C L E I N F O

Keywords:
Driver state
Driving environment
Sleepiness
Fatigue
Machine learning
Physiological state
Sleep deprivation

A B S T R A C T

Conditionally automated vehicles (Level 3 SAE) are emerging on the roads, but long periods without engaging in a non-driving-related task can reduce drivers' vigilance. This study aims to determine whether driver fatigue can be accurately predicted using physiological signals and machine learning (ML) techniques in such a context. 63 young drivers completed two separate conditional automated drives of 30 min each, in either a rural or a urban area. Half of them had been mildly sleep-deprived the previous night (slept less than six hours). Electrocardiogram (ECG), electrodermal activity (EDA), and respiration were collected, along with subjective measures of sleepiness and affective state. Using ML, sleep deprivation, driving environment, and sleepiness could be predicted from physiological features with an accuracy of 99%, 85%, and 73% respectively. Signal segmentation increased model accuracy, and EDA features were the most predictive. The differences between the results obtained from statistical analyses of sleepiness measures and the accuracy achieved by ML models are discussed. The results of this empirical study indicate that even mild sleep deprivation affects the physiological state of drivers, which can have serious consequences when combined with long periods of inactivity. Car manufacturers and researchers should take this into account when designing intelligent systems capable of providing drivers with appropriate warnings before a critical situation arises.

Introduction

Despite impressive technological advancements in the past decade, driving is still highly dangerous. In 2019 alone, there were more than 20,000 traffic accidents in Switzerland, of which more than 9,000 involved passenger vehicles (Swiss Federal Statistical Office, 2020). Impaired driving skills, such as distraction, alcohol, or fatigue, are the cause of 25 % of the victims of serious accidents each year (Hertach et al., 2020). To tackle this, more and more of the driver's tasks are being assisted or taken over by automation to increase road safety.

The Society of Automotive Engineers (SAE) created a taxonomy to designate the level of vehicle automation (SAE, 2018). In this manuscript, L*X*−SAE defines the automation level according to this taxonomy. X ranges from 0 to 5, the higher the automation level, the more tasks are handled by the vehicle. L2-SAE vehicles are currently on the roads while L3-SAE ones are emerging. At L3-SAE, drivers are no longer responsible for driving the vehicle and monitoring its environment. They could therefore engage in non-driving-related tasks, but be ready to react if a takeover is required. Several factors can impair the takeover performance, such as fatigue, sleepiness, or distraction due to the engagement in a non-driving-related task (Jarosch et al., 2019; Nakajima and Tanaka, 2017; Wandtner et al., 2018). A continuous and non-intrusive assessment of the driver's state is thus needed for L3-SAE vehicles to operate safely and avoid accidents.

Previous research has shown that the above-mentioned factors can be detected accurately from physiological data using machine learning (ML) techniques (Healey and Picard, 2005; Patel et al., 2011; Chen et al., 2017; Darzi et al., 2018; Meteier et al., 2021a; Meteier et al., 2021b). Recent advances in wearable sensors make physiological signals a credible candidate for measuring driver state in future vehicles (Meteier et al., 2022). Although not yet widely implemented in vehicles on the market, research is needed to prove the potential of physiological data to

* Corresponding author.
E-mail address: quentin.meteier@hes-so.ch (Q. Meteier).

https://doi.org/10.1016/j.trip.2024.101148
Received 5 October 2023; Received in revised form 23 February 2024; Accepted 18 June 2024
Available online 21 June 2024

robustly and accurately assess driver state at high levels of automated driving (L3-SAE and above). The main aim of this study is to fill research gaps by examining how accurately physiological activation related to sleep deprivation, driving environment, and sleepiness can be detected using ML techniques, specifically at L3-SAE.

Literature review

Theoretical constructs

Fatigue and sleepiness

In 2020, fatigue was among the five most frequent causes of accidents in Switzerland (Swiss Federal Statistical Office, 2020). Special attention must be paid when talking about fatigue, as it is a complex concept and several definitions can be found in the literature. It is close to, but different from, sleepiness, and people are often tempted to use the two terms interchangeably (Shen et al., 2006). To focus on fatigue in the context of driving, May and Baldwin (2009) have defined a model of driver fatigue distinguishing three types of fatigue, with their causes, consequences, and interactions. The three types of fatigue are sleep-related, passive task-related, and active task-related fatigue. Only sleep-related and passive task-related fatigue will be addressed in this article.

The sleep-related fatigue can be caused by several factors such as sleep deprivation, duration of wakefulness, or circadian rhythm. The latter is the individual's endogenous rhythm (Potter et al., 2016) regulating the sleepiness throughout the day (Valdez, 2019). For instance, the circadian rhythm is responsible for individuals being generally less alert during night hours (Costa, 1996; Moller et al., 2003; Reimer et al., 2007; Garde et al., 2020). Sleep-related fatigue consequently affects individuals' alertness (Jewett et al., 1999; Philip et al., 2003), which can be critical for drivers, especially if they need to react quickly after a takeover request at L3-SAE driving.

While automated driving systems are increasingly making driving easier, they also require less and less human interaction. This places the driver outside the control loop of the vehicle, which can lead to a decrease in attention and situational awareness (Heikoop et al., 2019), and consequently increase reaction time when a takeover is requested (Jarosch et al., 2019). In particular, the second type of fatigue defined by May and Baldwin (2009), the passive task-related fatigue, refers to this issue. This type of fatigue can also be described as sleepiness. In this article, the term sleepiness is used to refer to passive task-related fatigue. It typically occurs during a monotonous driving phase, under conditions of mental underload, or during periods of automated driving (May and Baldwin, 2009). Sleepiness might be a serious problem for driving safety in automated systems, as drivers may not be involved in the driving task for long periods. Thus, their engagement and alertness might be reduced (Radlmayr et al., 2018). Previous studies found that drivers of partially (L2-SAE) and highly (L4-SAE) automated vehicles had higher levels of sleepiness than drivers of manual vehicles, particularly in monotonous environments (Ahlström et al., 2021; Schömig et al., 2015). In addition, sleepiness may exacerbate sleep-related fatigue, with both leading to a decreased driving performance (May and Baldwin, 2009), increasing the risk of accidents (Cai et al., 2021; Guo et al., 2021). Therefore, it is necessary to assess sleep-related fatigue and sleepiness to reduce the risk of accidents. For that, physiological measures can be used to assess driver fatigue over time, both in manual and automated driving (Feldhütter et al., 2018; Zhou et al., 2020).

Fatigue can also alter the driver's affect (Palmer and Alfano, 2017). According to Russel's circumplex model (Russell, 1980), affect can be described with the two dimensions arousal and valence. Valence is the level of pleasantness on a scale from negative to positive, and arousal is the intensity of activation, ranging from low to high (Bestelmeyer et al., 2017). It has been shown, that loss of sleep is related to negative affect (Tomaso et al., 2020), which in turn impacts driving performance, leading to impaired takeover quality (Du et al., 2020) as well as more frequent errors (Rowden et al., 2011).

Driving environment

Driving is a cognitively and visually demanding activity. The level of demand depends on different aspects such as interpersonal differences as well as contextual factors (Shinar, 1978; Underwood, 2007; Llaneras et al., 2017; di Flumeri et al., 2018). The visual workload induced by the driving environment is one such contextual factor, resulting from the complexity of the environment, i.e., cars, traffic signs, road type, and landscape (Lyu et al., 2017). These aspects can vary considerably depending on the driving situation. In automated driving, a mental overload could consequently impair the takeover quality (Scharfe et al., 2020). The driving environment can also have an impact on subjective measures of driver's affect. For instance, cognitively demanding traffic incidents, such as high traffic density (Zepf et al., 2019), can induce higher negative emotions. Also, congestion is known for eliciting negative emotions, such as aggression or frustration (Shinar and Compton, 2004). As mentioned above, negative emotions are a risk factor for road safety, even in automated driving.

Similarly to complex driving scenarios, the monotonous nature of roads leads to a decrease in driving performance, probably due to a state of mental underload (Larue et al., 2011; Thiffault and Bergeron, 2003a; Thiffault and Bergeron, 2003b). Therefore, the environment is an important determinant of driving performance and should also be considered in conditionally automated driving.

Prediction of driver state using physiological signals and machine learning

Fatigue and sleepiness

Previous studies that predicted driver fatigue using ML techniques and at least an electrocardiogram (ECG), electrodermal activity (EDA) or respiration (RESP) were reviewed (Awais et al., 2017; Bundele and Banerjee, 2009; Darzi et al., 2018; Fujiwara et al., 2019; Kiashari et al., 2020; Kundinger et al., 2020; Kundinger and Riener, 2020; Lee et al., 2019; Li and Chung, 2013; Patel et al., 2011; Rigas et al., 2011; Sharma and Bundele, 2015; Vicente et al., 2011; Wang et al., 2017). Most of them were conducted in a driving simulator at L0-SAE (manual driving). To induce sleepiness, 10 to 30 participants were asked to drive manually for a long period (20 min to 6 h) on a highway, in a monotonous environment without traffic, at a reasonable speed (between 80 and 110 km/h). The manipulation of sleepiness was often combined with sleep-related fatigue to enhance global fatigue (Darzi et al., 2018; Kiashari et al., 2020; Patel et al., 2011; Vicente et al., 2011). Some other studies were conducted at times of low alertness (e.g., after lunch or at night) (Fujiwara et al., 2019; Lee et al., 2019; Li and Chung, 2013).

Heart-rate variability (HRV) features were used in almost all studies. There is no consensus on the time window used to calculate the physiological indicators, as it ranged from 2 s (Bundele and Banerjee, 2009), to 1–2 min (Vicente et al., 2011; Li and Chung, 2013; Lee et al., 2019; Kiashari et al., 2020) to 5–15 min (Fujiwara et al., 2019; Kundinger et al., 2020; Kundinger and Riener, 2020; Rigas et al., 2011). Baseline correction was rarely applied to features (Vicente et al., 2011; Darzi et al., 2018).

Most studies carried a binary classification task to predict sleepiness (alert vs. drowsy) using a between-subject evaluation approach. Self-reported measures or external observations were often used as ground truth. Wang and colleagues achieved 99 %-accuracy in predicting fatigue induced by extended driving duration from EDA, RESP and pulse oximetry features (Wang et al., 2017). At the L2-SAE level, 99 % and 98 % accuracy was achieved to distinguish drivers' fatigue at two and three different levels. Two studies achieved 90 %-accuracy in predicting a combination of sleep-related fatigue and sleepiness (Kiashari et al., 2020; Fujiwara et al., 2019). However, no reviewed studies predicted sleep-related fatigue alone.

Specifically at L3-SAE, the only study that predicted driver sleepiness using ML techniques and physiological data achieved high accuracy (adjusted R2 score of 0.996) (Zhou et al., 2021). The percentage of eye closure (PERCLOS) was used as the ground truth, but driving-related features were used as input data even though it would not be possible to use them under real driving conditions. All these previous studies reveal a research gap in predicting sleep-related fatigue and sleepiness separately, specifically at L3-SAE, and only from physiological signals.

Driving environment

Driving in a variety of environments can affect the drivers' physiological state and behaviour, which can be detected with ML techniques. Healey and Picard (2005) achieved 97.4 %-accuracy to distinguish three levels of stress (low/rest, medium/highway, high/town) induced in real driving condi- tions from five-minute windows of physiological data. Chen et al. (2017) even achieved 99.9 %-accuracy on the same dataset with a leave-one-out within-subject validation approach, while 89.7 %-accuracy was reached with a between-subject approach. In a moving-base driving simulator, Darzi et al. (2018) showed that the physiological activation related to the driving environment (highway vs. town) could be predicted with 86.8 %-accuracy only from physiological signals using a k-fold cross-validation approach. 83.3 %- accuracy was achieved only with vehicle kinematics, and 91.3 % by combining both sources of data.

The present study

As reported above, ML has been widely used to detect a change of physiological activation due to fatigue or driving environment from physiological signals in manual driving (L0-SAE) (Healey and Picard, 2005; Chen et al., 2017; Darzi et al., 2018; Patel et al., 2011) or at the L2-SAE for fatigue (Kundinger and Riener, 2020; Kundinger et al., 2020). However, no study was carried out to detect these states at the L3-SAE level using only physiological data. Similar studies were conducted at L3-SAE but rather aimed at predicting the driver's mental workload induced by non-driving-related tasks (Meteier et al., 2021a; Meteier et al., 2021b).

To fill this research gap, the goal of this driving simulator study is to investigate whether the physiological change related to sleep deprivation, sleepiness, and driving environment can be accurately detected using ML techniques specifically at L3-SAE. Self-reported measures of sleepiness and affect were also collected before and after each drive.

Based on previous studies, the following hypotheses were formulated to address the research questions:

- (H1): Sleep deprivation
  - (a) Sleep deprivation has a negative effect on self-reported sleepiness, even after the drive (Philip et al., 2003, 2005).
  - (b) Sleep deprivation impairs drivers' affect (Pires et al., 2016; Tomaso et al., 2020).
  - (c) Sleep deprivation significantly impairs the physiological state of the driver (Vicente et al., 2011; Fujiwara et al., 2019; Altemus et al., 2001; Liu et al., 2015; Rault et al., 2019; Ahlström et al., 2021).
  - (d) No hypothesis can be formulated for predicting sleep deprivation with ML, as no previous study has addressed this issue.
- (H2): Duration of driving scenario
  - (a) Self-reported sleepiness will increase over the course of the L3-SAE drive – i.e., higher before the takeover than before the drive (Ahlström et al., 2021; de Naurois et al., 2017; Vogelpohl et al., 2019).
  - (b) Drivers' affect should be altered due to the boring nature of the monotonous driving task. Arousal should decrease due to lack of engagement, which could frustrate them and thus also alter valence (van Hooft and van Hooff, 2018).
  - (c) Similar accuracy (between 95 % and 100 %) than previous studies conducted at L2-SAE should be achieved for the prediction of sleepiness (before the drive vs. before the takeover) (Kundinger and Riener, 2020; Kundinger et al., 2020). This is because sleepiness might occur more quickly at L3-SAE, the driver being passive at the wheel.
- (H3): Driving environment
  - (a) Self-reported sleepiness should be higher in a monotonous driving environment (Thiffault and Bergeron, 2003a; Thiffault and Bergeron, 2003b).
  - (b) The complexity of the driving environment impairs driver's affect (Zepf et al., 2019; Shinar and Compton, 2004).
  - (c) The complexity of the driving environment has a negative effect on the physiological state (Healey and Picard, 2005; Chen et al., 2017).
  - (d) For the prediction with ML techniques, the accuracy achieved at L3-SAE should be lower than in previous studies because the driver was actively driving, either in real-driving situation (Chen et al., 2017; Healey and Picard, 2005) or in a driving simulator (Darzi et al., 2018).

Material and methods

Participants

63 young participants (M = 23.76, SD = 7.24) took part in this experiment, including 45 women and 18 men. 15 of them had consumed caffeine in the 12 h preceding the experiment. The majority (n = 60) reported eating between a small portion and a normal portion of food at breakfast.

On average, the subjects had held a driving license for 5 years (SD = 7.31) and reported driving 5791 km per year (SD = 7704.00 km). Most of them drove once or several times a week (n = 41) and had not had an accident in the last three years (n = 56). Only 12 subjects had prior experience of the L2-SAE automation level. Seven of them had previously participated in a driving simulator study. All participants gave written consent.

Most of the participants usually slept between 11 pm and midnight (n = 41) and woke up between 7 am and 8 am (n = 38) during the previous month. Thus, most (n = 42) were accustomed to sleeping between 7 and 8 h per night, which is consistent with previous research on average student sleep time (Ackermann et al., 2015; Borisenkov et al., 2010; Gilbert and Weaver, 2010; Ness and Saksvik-Lehouillier, 2018; Okano et al., 2019). The majority said it generally took them between 10 and 20 min to fall asleep.

Experimental design

The study had a 2x2x2 mixed design. Sleep deprivation was the first two-level between-subjects factor, with each level corresponding to the amount of sleep requested the night before the experiment (sleep deprived vs. normal sleep). Participants were instructed to sleep either less than six hours or at least seven hours. The minimum duration was set at seven hours for the control group because students sleep an average of seven to eight hours per night according to various studies (Ackermann et al., 2015; Borisenkov et al., 2010; Gilbert and Weaver, 2010; Ness and Saksvik-Lehouillier, 2018; Okano et al., 2019). Such sleep deprivation can be considered acute because it is mild, punctual, and recent (only the night before the experiment) (Shen et al., 2006).

The driving environment was a two-level within-subjects factor (rural vs. urban). Both environments differed in terms of traffic, road variability, and scenery. The scenario order was the second between-subjects factor (rural first vs. urban first), with half of the individuals driving first in the urban environment and half in the rural one. The experiment time was controlled (10:00 am or 4:00 pm) to avoid an effect of the circadian rhythm (Valdez, 2019).

Material and instruments

Driving simulator

Fig. 1 shows a participant in the fixed-base driving simulator during the training phase. It includes a driver's seat, a Logitech steering wheel with the gas pedal, brake, and clutch pedals. A television screen (65 in.) displayed the driving scenario. Another screen (13.1 in.) located behind the steering wheel displayed the vehicle's dashboard with the autopilot mode (on/off/takeover), the car's speed, and the number of engine revolutions per minute. A computer ran the driving simulation and recorded the driving data. A psychomotor vigilance task was also implemented (pressing a button when a red dot appeared every five minutes on the screen). Results from this task are not detailed in this manuscript.

Driving environments

Two different environments were designed to study their effect on the driver's state. Both were developed with the Unity software. The first one was a monotonous rural environment. A screenshot of the scenario can be seen in Fig. 2a. There was no traffic, crosswalks, or traffic lights, and the road was slightly curvy. The roadside was lined with trees and rocks that were of similar look and shape. The second scenario was an urban environment (see Fig. 2b). The city had medium-density traffic and the road had many intersections with and without traffic lights, where the car had to turn or stop. The roadside was varied, with pedestrians at some intersections, bike racks, parked cars, trees, parks, large buildings with different facades, bridges, and street lights. Conditionally automated driving functions were implemented for both scenarios. Both lasted 25 to 30 min, which should be sufficient for sleepiness to occur (Thiffault and Bergeron, 2003b; de Naurois et al., 2017; Vogelpohl et al., 2019). Each scenario ended with a takeover request (TOR) triggered by a dog crossing the road. The route taken by the vehicle was the same for all participants. The time spent in the urban scenario varied slightly between participants due to the random activation of traffic lights and traffic.

Fig. 1. A participant in the driving simulator.

Other material

BioPac MP36 hardware with the BioPac Student Lab 3.7.7. software were used to collect physiological signals from the drivers on a computer at a sampling rate of 1000 Hz. Pre-gelled disposable Ag/AgCl electrodes (EL507 and EL503, Biopac) connected to wire sets (SS57LA and SS2LB, Biopac) collected EDA and ECG signals from the subjects. Two electrodes placed at the tip of the ring and middle fingers of the non-dominant hand recorded EDA. Three electrodes recorded ECG, one above each malleolus and one on the right forearm. A breathing belt (SS5LB, Biopac) collected the breathing signal via chest expansion. A tablet displayed the questionnaires created on Unipark [1].

Measures

Participants' EDA, ECG, and RESP were recorded throughout the experiment. A wide range of indicators was calculated from these signals (see Section 3.7.2 and Table 9). The mean tonic EDA level, frequency of skin con– ductance responses (number of occurrences per minute), the ratio of power in low and high-frequency bands of HRV (LF/HF), and estimates of Respiratory Sinus Arrhythmia (RSA) using the Gates method (Gates et al., 2015) were selected as measures of physiological activation that may be influenced by sleep deprivation and driving environment (Posada-Quintero et al., 2017; Liu et al., 2015; Burton et al., 2010; Chua et al., 2012; Schmitt et al., 2015; Katona and Jih, 1975; Healey and Picard, 2005; Darzi et al., 2018).

Fig. 3 shows the content of each questionnaire administered during the experiment. The modified version of the Karolinska Sleepiness Scale (KSS) (Åkerstedt and Gillberg, 1990; Kaida et al., 2007) was used to measure the driver's self-rated sleepiness on a scale of 1 (extremely alert) to 10 (extremely [1]https://www.unipark.com/en/.

sleepy, cannot stay awake). Valence and arousal were also assessed using the animated version of the Self-Assessment Manikin (AniSAM) (Sonderegger et al., 2016). Other measures were collected (takeover behavior, situation awareness, vigilance, trust user experience) were also collected but are not analysed in this manuscript.

Experimental procedure

An overview of the experimental procedure, is shown in Fig. 3. On the day of the experiment, the participants were first informed about the context of the study and the course of the experiment. After signing the consent form, they had to fill in questionnaire 1 (see Fig. 3). Then, the electrodes and the breathing belt were installed on the participant to record the physiological signals.

The experiment started with a five-minute phase to record the physiological baseline. The car was driving at L3-SAE, while the participants observed the environment. Then, they could drive manually the simulator to be familiar with the commands and the takeover process (*Training* phase). The experimenter ensured that the participant was able to take control appropriately before starting the main scenarios.

In the latter, drivers were instructed to monitor the environment and take-over control only when requested by the car, by turning the wheel, braking, or pressing a button on the steering wheel. In addition, they had to perform a psychomotor vigilance task, which consisted of pressing a button on the steering wheel every five minutes when a red dot appeared on the screen. After the TOR, participants were asked to stop the vehicle and complete the questionnaire on the tablet (questionnaire 2 or 3, see Fig. 3). After the second scenario, the electrodes were removed. To thank the participants, all received chocolate and the university psychology students were credited for the hours spent as test subjects.

(a) Rural scenario.
(b) Urban scenario.

Fig. 2. The two driving scenarios.

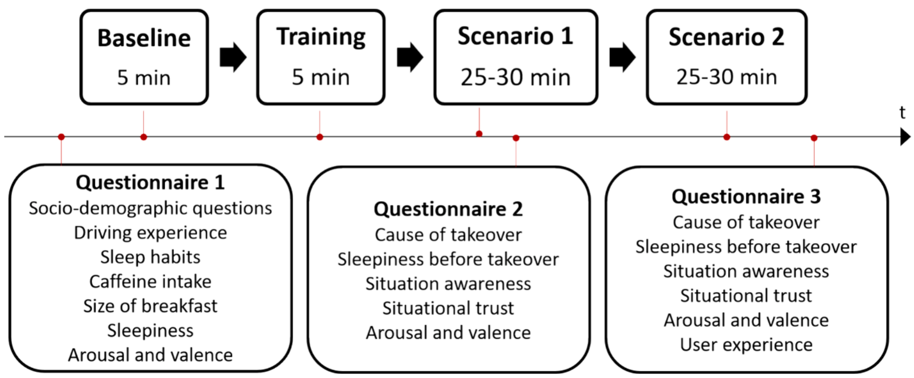

Fig. 3. Overview of the experimental procedure.

Statistical analysis

An independent samples *t*-test controlled for the experimental manipulation of sleep deprivation using the subjective measure of sleepiness reported before the experiment (KSS of questionnaire 1).

To investigate the effect of *sleep deprivation* (sleep deprived vs. normal sleep) and *scenario order* (urban first vs. rural first) as between-subjects factors, the effect of *driving environment* (rural vs. urban) as within-subjects factor, and the interaction effects, a mixed repeated measures analysis of variance with the baseline measure as covariate (ANCOVA) was performed on physiological measures.

To investigate the change of self-reported sleepiness (KSS), arousal, and valence over time, the *time* factor (before vs. after) was included as a within-subjects factor. For sleepiness, the second measure corresponded to sleepiness before the takeover, while for arousal and valence, the measure corresponded to their state after the takeover. Measures after the first scenario were used as the pre-drive measures of the second scenario.

For all analyses, if Mauchly's test indicated that the assumption of sphericity was violated ($p < 0.05$), Greenhouse-Geiser sphericity corrections were applied. Post-hoc tests with Bonferroni correction were done when the effect of time was significant.

Classification of the Drivers' state

The classification tasks

The following classification tasks were performed in this work:

- Classification task 1: Prediction of sleep deprivation condition (sleep deprived vs. Normal sleep). Drivers' physiological features calculated in both environments were used
- Classification task 2: Prediction of the driving environment (urban vs. rural) in which participants were driving. Drivers' physiological features calculated in both environments were used.
- Classification task 3: Prediction of drivers' sleepiness (beginning of the drive vs. Before the takeover). Only drivers' physiological features calculated in the rural scenario were used

For all classification tasks, the effect of sensor fusion and segmentation level on model performance was tested. The segmentation consisted of splitting the raw signals into multiple segments and calculating the physiological indicators for each segment. Fig. 4 shows an overview of this process.

For classification task 3, the goal was to predict whether a given

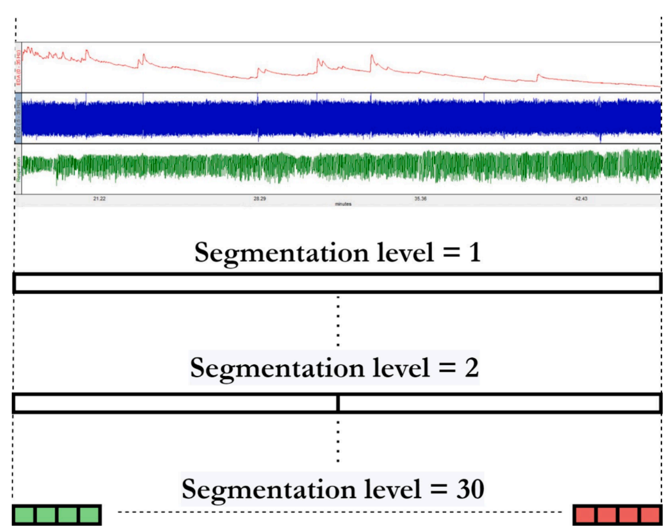

Fig. 4. Procedure of segmentation for the classification tasks. For task 3, the first segments (in green) were labeled as *low sleepiness* and the last segments before the takeover situation were labeled as *high sleepiness*. Red signal: EDA, blue signal: ECG, green signal: RESP. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

segment corresponded to the beginning (low sleepiness) or to the end (high sleepiness) of the drive, before the TOR. Table 1 presents the different numbers of segments tested for each segmentation level.

The classification pipeline

A pipeline was implemented in Python to perform the three tasks described above. Physiological signals were processed using the Neurokit library (Makowski et al., 2021). To account for the physiological state of the participants before starting the driving experiment (baseline phase, see Fig. 3), 2 features were calculated for each physiological indicator:

- Dr: The value of the physiological indicator during the considered time window while driving
- Dr-Bl: The difference between the value of the physiological indicator during the considered time window while driving and the value of the same indicator during the baseline phase (ΔDriving-Baseline).

For each segmentation level (1, 2, 5, 10, 20, and 30) and each scenario, 224 features corresponding to 112 physiological indicators (10 from EDA, 74 from ECG, 21 from RESP, 7 from RSA) were calculated. Table 9 presents indicators calculated from each physiological signal. Features were normalised between the first quartile and third quartile of the data distribution (RobustScaler method (Pedregosa et al., 2011)). The top 20 features were selected based on univariate statistical tests of ANOVA (SelectKBest method (Pedregosa et al., 2011)), which were used to train three different ML algorithms: a random forest (RF), a neural network with one hidden layer (NN), and a k-nearest neighbors (KNN). They were implemented with the scikit learn framework (Pedregosa et al., 2011).

A 5-times 4-fold cross-validation procedure was used to train and evaluate the ML models. At each iteration, the data set was randomly divided into a training set (80 %) and a test set (20 %). A 4-fold cross-validation with a grid search approach was employed on the training set (Mosteller and Tukey, 1968) to tune hyperparameters. The best model on the 4 folds was then evaluated on the test set. The mean weighted F1-score achieved by the 3 ML models over the 5 iterations is reported. In addition, the Shapley Additive exPlanations (SHAP) module (Lundberg and Lee, 2017) was used to return the 10 most predictive features for each classification task.

Results

Statistical analysis

Except for the manipulation check, all the results from the statistical analysis are reported in Table 2, for all dependent measures.

Manipulation check

The statistical analysis showed a significant effect of mild sleep deprivation on sleepiness before the experiment, with the sleep-deprived drivers (M = 5.68, SD = 1.42) feeling more drowsy than the control group (M = 4.03, SD = 1.60; t(61) = -4.31, $p < 0.001$, d = -1.09). No significant effect of experiment time was found on sleepiness before the experiment (t(61) = 0.63, $p > 0.05$, d = 0.16).

Table 1
The number of segments tested for classification task 3.

Segmentation level	Nb. of segments
5	1
10	1 to 3
20	1 to 6
30	1 to 10

Table 2
Results of the statistical analysis, reported following this format: F-value (η^2). Significant effects are shown in bold, with the corresponding p-value: * $p < 0.05$, ** $p < 0.01$,

Measures	SD	DE	SO	DE x SO	Time	DE x Time
Sleepiness	**14.35 (0.20)** ***	1.59 (0.03)	3.21 (0.05)	**4.79 (0.07)***	**8.29 (0.12)** **	**8.11 (0.12)** **
Arousal	F < 1	2.28 (0.04)	1.43 (0.02)	3.53 (0.06)	3.53 (0.06)	**4.88 (0.08)***
Valence	1,59 (0.03)	F < 1	F < 1	**6.45 (0.10)***	**6.45 (0.10)***	F < 1
Tonic EDA	F < 1	F < 1	2.95 (0.05)	**7.23 (0.12)***	/	/
NS-SCRs	F < 1	F < 1	1.85 (0.03)	F < 1	/	/
SDNN	F < 1	1.10 (0.02)	1.84 (0.03)	**30.01 (0.36)** ***	/	/
LF/HF	F < 1	F < 1	F < 1	F < 1	/	/
RSA	F < 1	F < 1	**7.26 (0.12)** **	**27.19 (0.33)** ***	/	/

*** $p < 0.001$. Interaction effects for which there were no significant effects on any measures are not displayed in this table. *SD = Sleep Deprivation; DE = Driving Environment; SO.*
= Scenario Order.

Sleepiness

A significant effect of sleep deprivation and time was found on sleepiness (see Table 2), which ***validates (H1a) and (H2a)***. During the experiment, sleep-deprived drivers felt sleepier (M = 6.04, SE = 0.24) than those who slept normally (M = 4.77, SE = 0.24). Participants also reported being sleepier after 30 min of conditionally automated driving (M = 5.61, SE = 0.18) than before (M = 5.20, SE = 0.18). In addition, a significant interaction effect of driving environment and time was found (see Table 2). In the rural environment, participants reported being sleepier after the drive (before the takeover; M = 5.95, SE = 0.21) than before (M = 5.05, SE = 0.21, t(59) =−4.04, $p < 0.001$). However, they did not report a change of sleepiness before (M = 5.34, SE = 0.21) and after driving (before the takeover; M = 5.27, SE = 0.21, t(59) = 0.30, $p > 0.05$) in the urban one. This suggests that ***(H3a) is validated***, but this is further discussed in section 5.4.4.

Besides, a significant interaction effect of the driving environment and scenario order was found (see Table 2). Analysis of simple main effects revealed that sleepiness was significantly higher in the rural environment than in the urban one for participants who drove first in the urban scenario (F(1, 54) = 5.93, $p < 0.05$), but there was no difference in sleepiness between both environments for those who drove first in the rural scenario (F(1, 54) = 0.434, $p > 0.05$).

Arousal and valence

A significant interaction effect of driving environment and time was found on arousal (see Table 2). In the rural environment, participants reported to be less aroused after the drive (M = 2.14, SE = 0.09) than before (M = 2.44, SE = 0.09, t(59) = 2.77, $p < 0.05$), ***which validates (H2b)***. However, they did not report a change of arousal between before (M = 2.29, SE = 0.09) and after driving (M = 2.43, SE = 0.09, t(59) = -1.30, $p > 0.05$) in the urban one.

A significant effect of time was found on valence (see Table 2). Participants reported lower scores of valence after the drive (M = 3.49, SE = 0.08) than before (M = 3.61, SE = 0.08). ***This also validates (H2b)***. Besides, a significant interaction effect of driving environment and scenario order was found on valence (see Table 2). Nevertheless, planned comparisons did not show any significant differences between conditions ($p > 0.05$).

Other single and interaction effects were not significant, so ***(H1b) and (H3b) are refuted.***.

Physiological state of drivers

All physiological indicators measured during baseline showed a significant effect as covariate, including the tonic EDA level ($F(1, 54) = 152.53$, $p < 0.001$, $\eta^2 = 0.71$), frequency of skin conductance responses ($F(1, 54) = 40.53$, $p < 0.001$, $\eta^2 = 0.31$), LF/HF ($F(1, 54) = 83.67$, $p < 0.001$, $\eta^2 = 0.55$), SDNN.

pp ($F(1, 54) = 117.51$, $p < 0.001$, $\eta^2 = 0.58$), and RSA ($F(1, 54) = 411.60$, $p < 0.001$, $\eta^2 = 0.82$).

A significant interaction effect of the driving environment and scenario order was found on tonic EDA level (see Table 2), which can be seen in Fig. 5. Analysis of simple main effects revealed that tonic EDA level was significantly different between rural and urban scenarios for participants who drove first in the urban scenario ($F(1, 54) = 4.69$, $p < 0.05$), but not for those who drove first in the rural scenario ($F(1, 54) = 2.79$, $p > 0.05$). Apart from these significant effects, no other significant effect was found on both EDA measures ($p > 0.05$).

Regarding HRV measures, a significant interaction effect of the driving environment and scenario order was found on SDNN (see Table 2), which is shown in Fig. 6. Post-hoc comparisons revealed that for participants who drove first in the urban scenario, SDNN was higher in the rural scenario afterwards ($t(54) = 3.01$, $p < 0.05$), while for participants who drove first in the rural scenario, SDNN was higher in the urban scenario ($t(54) = -4.72$, $p < 0.001$). Otherwise, no other significant effect was found on HRV measures ($p > 0.05$).

A significant effect of scenario order was found on RSA (see Table 2). RSA was higher for participants who drove first in the rural environment ($M = 15.12$, $SE = 0.34$) than those who drove in the urban environment first ($M = 15.04$, $SE = 0.34$). In addition, a significant interaction effect of driving environment and scenario order was found on RSA (see Table 2), which can be seen in Fig. 7. Post-hoc comparisons revealed that for participants who drove first in the urban scenario, RSA was higher in the rural scenario afterwards ($t(54) = 3.15$, $p < 0.05$), while for participants who drove first in the rural scenario, RSA was higher in the urban scenario ($t(54) = -4.21$, $p < 0.001$).

No other significant effect was found on RSA ($p > 0.05$), including the sleep deprivation and the driving environment alone. So ***(H1c) and (H3c) are refuted.***

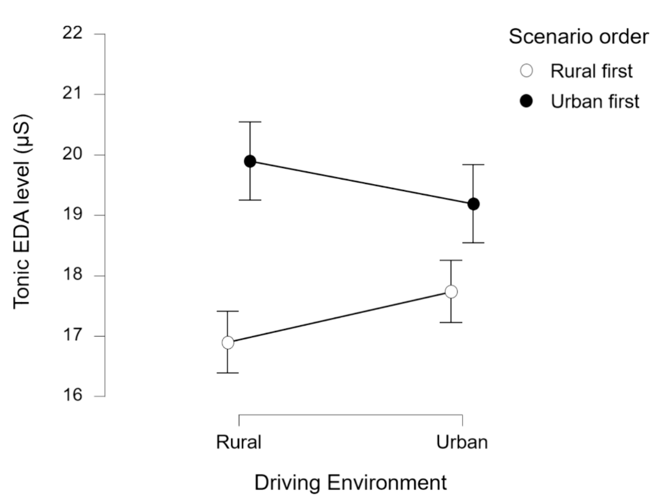

Fig. 5. Tonic EDA level (in μS) as a function of driving environment and scenario order. *Error bars represent the 95% confidence interval.*

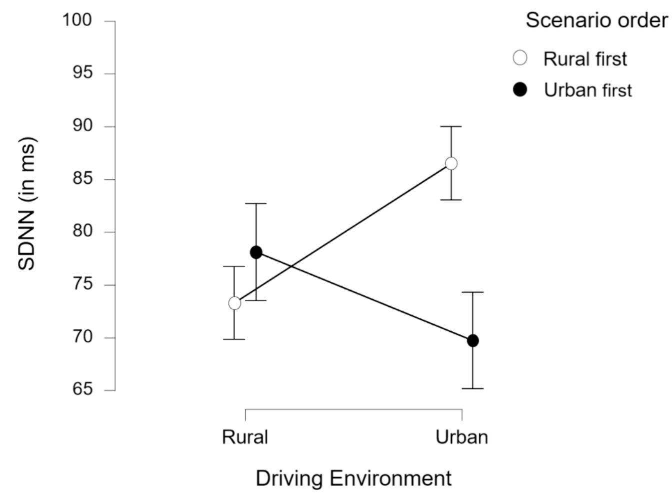

Fig. 6. SDNN (in milliseconds) as a function of driving environment and scenario order. *Error bars represent the 95% confidence interval.*

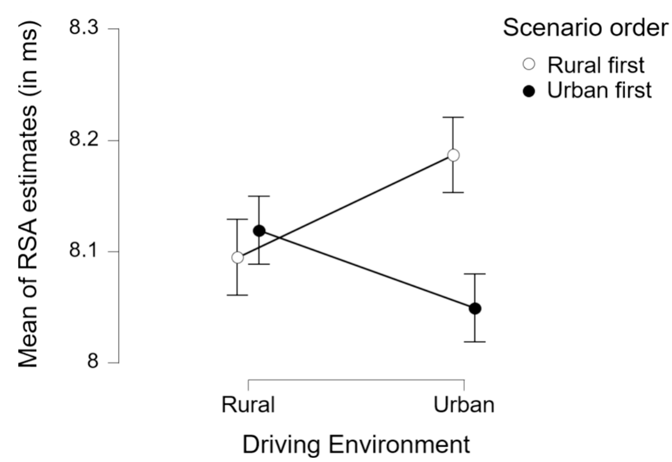

Fig. 7. Mean of RSA estimates (in milliseconds) calculated with the Gates method, as a function of driving environment and scenario order. *Error bars represent the 95% confidence interval.*

Classification of drivers' state using machine learning

Classification tasks 1 and 2

For each combination of signal(s), Table 3 and 4 shows the average F1-score (with standard deviation) achieved by ML models to predict drivers' conditions (sleep deprivation and driving environment) over the 5 iterations of the classification procedure. A segmentation level of 1 was used, meaning that features used as inputs of ML models were computed from the entire signals (from the beginning of the scenario

Table 3
Effect of sensor fusion on average F1-score achieved by the model for classification task 1 (sleep deprivation). *Bold values indicate the best score achieved across all combinations of signals.*

Selected signal(s)	Classifier	F1-score (Mean (SD))
EDA	NN	0.66 (0.10)
ECG	NN	0.69 (0.10)
RESP	NN	0.70 (0.07)
EDA + ECG	**NN**	**0.71 (0.06)**
ECG + RESP	NN	0.68 (0.14)
EDA + RESP	NN	0.71 (0.10)
EDA + ECG + RESP	NN	0.70 (0.12)

Table 4
Effect of sensor fusion on average F1-score achieved by the model for classification task 2 (driving environment). *Bold values indicate the best score achieved across all combinations of signals.*

Selected signal(s)	Classifier	F1-score (Mean (SD))
EDA	RF	0.41 (0.06)
ECG	RF	0.49 (0.04)
RESP	RF	0.54 (0.07)
EDA + ECG	KNN	0.51 (0.16)
EDA + RESP	**NN**	**0.55 (0.08)**
ECG + RESP	**KNN**	**0.55 (0.08)**
EDA + ECG + RESP	RF	0.50 (0.10)

until the takeover request). The *Classifier* column reveals the ML algorithm that achieved the best accuracy. Besides, Tables 5 and 6 show the effect of segmentation (from 1 to 30) on the accuracy of ML models for predicting drivers' conditions (sleep deprivation and driving environment). For each segmentation level, the best average F1-score (with standard deviation) achieved over the 5 iterations of the procedure is reported. In each case, the ML algorithm that achieved it, the input signals, and the number of samples used are reported. As the highest score achieved for the prediction of the driving environment is an F1-score of 0.81, ***(H3d) is validated*** because it is lower than previous similar research. This is discussed in section 5.4.3.

Classification task 3

Table 7 reports the best average F1-score (with standard deviation) achieved over the 5 iterations of the classification procedure. In each case, the ML algorithm that achieved it, the input signals, and the number of segments used are reported. The best score achieved is 0.69, which also ***refutes (H2c)***.

This is discussed in section 5.4.4.

Most predictive features

Table 8 shows the 10 most predictive features used to achieve the best accuracy for each classification task. Feature rankings were calculated using the SHAP library once the best model had been trained and evaluated on the test set.

Discussion

Manipulation check

Regarding the manipulation of sleep deprivation, a significant difference was found in self-reported sleepiness before starting the experiment. The sleep-deprived group felt sleepier than the control group. These results validate the experimental design, as the aim was to test participants at times of similar alertness and to study the effect of mild sleep deprivation on the drivers' state.

Table 5
Effect of segmentation on average F1-score achieved by the ML models for classification task 1 (sleep deprivation). *Bold values indicate the best score achieved across all segmentation levels. F1-score = Mean (standard deviation). Segm. lvl = Segmentation level.*

Segm. lvl	Signal(s)	Classifier	F1-score	Samples
1	EDA + RESP	NN	0.71 (0.10)	120
2	EDA + RESP	RF	0.84 (0.06)	240
5	EDA	NN	0.88 (0.03)	599
10	EDA	NN	0.95 (0.02)	1190
20	EDA	NN	0.98 (0.01)	2377
30	**EDA**	**NN**	**0.99 (0.00)**	**3495**

Table 6
Effect of segmentation on average F1-score achieved by the ML models for classification task 2 (driving environment). *Bold values indicate the best score achieved across all segmentation levels. F1-score = Mean (standard deviation). Segm. lvl = Segmentation level.*

Segm. lvl	Signal(s)	Classifier	F1-score	Samples
1	EDA + RESP	NN	0.55 (0.08)	120
2	ECG + RESP	RF	0.66 (0.08)	240
5	EDA + ECG	NN	0.77 (0.02)	599
10	EDA + ECG	NN	0.75 (0.04)	1190
20	EDA	RF	0.81 (0.01)	2377
30	**EDA**	**RF**	**0.85 (0.01)**	**3495**

Table 7
Effect of segmentation on average F1-score achieved by the model for classification task 3 (sleepiness). *Bold values indicate the best score achieved across all segmentation levels. F1-score = Mean (standard deviation). Segm. lvl = Segmentation level. Nb. seg. = the number of segments for which the model achieved the best performance.*

Segm. lvl	Signal(s)	Nb. seg.	Classifier	F1-score
5	EDA + ECG + RESP	1	RF	0.56 (0.07)
10	ECG + RESP	3	RF	0.65 (0.04)
20	EDA + RESP	4	RF	0.69 (0.05)
30	**EDA**	**5**	**RF**	**0.73 (0.05)**

Table 8
The 10 most predictive features for each classification task (CT), sorted by ascending order. *EDA = Electrodermal activity, SCR = Skin Conductance Response, SD = Standard deviation, (Bl) = Corrected with baseline.*

Rank	CT1 – Sleep deprivation	CT2 – Driving environment	CT3 – Sleepiness
1	Mean Amplitude of SCRs	Minimum tonic EDA level	Maximum tonic EDA level (Bl)
2	Minimum tonic EDA level (Bl)	SD of raw EDA level	Mean raw EDA level (Bl)
3	SD of raw EDA level (Bl)	Maximum raw EDA level	SD of tonic EDA level
4	Number of SCRs per minute (Bl)	Minimum tonic EDA level (Bl)	Maximum raw EDA level (Bl)
5	Maximum tonic EDA level (Bl)	Maximum tonic EDA level	Number of SCRs per minute (Bl)
6	Number of SCRs per minute	Minimum raw EDA level (Bl)	Number of SCRs per minute
7	Minimum raw EDA level	SD of tonic EDA level	Maximum tonic EDA level
8	Mean Amplitude of SCRs (Bl)	SD of tonic EDA level (Bl)	Mean tonic EDA level (Bl)
9	Maximum raw EDA level (Bl)	Maximum raw EDA level (Bl)	Minimum tonic EDA level (Bl)
10	SD of tonic EDA level (Bl)	Number of SCRs per minute	Mean Amplitude of SCRs (Bl)

Sleepiness

The significant effect of sleep deprivation observed during the baseline phase was maintained throughout the rest of the experiment. Even though the sleep deprivation was mild, sleep-deprived drivers reported being sleepier during conditionally automated driving than participants who slept normally. ***Thus, (H1a) is validated.***

Regardless of their experimental condition, drivers felt sleepier after only 30 min of conditionally automated driving. ***Thus, (H2a) is validated.*** While they only had to observe the vehicle's environment and not perform the driving task, it confirms that the lack of engagement in a non-driving-related task can rapidly induce sleepiness. Data analysis revealed that drivers were particularly sleepier at the end of the rural scenario, but not after the urban one. ***We can thus argue that (H3a) is validated.*** This was mostly due to the monotonous nature of the rural

Table 9
Indicators calculated from raw physiological signals. Identical indicators computed from both ECG and respiration (RESP) signals are grouped together. *IBIs = interbeat intervals (ECG); BBs = breath-to-breath cycles (RESP). NS-SCRs = Non-Specific Skin Conductance Responses.*

Signal	Indicator	Domain	Description
EDA	Mean raw EDA level		The mean value of filtered EDA signal
	Min raw EDA value		The minimum value of filtered EDA signal
	Max raw EDA value		The maximum value of filtered EDA signal
	Std raw EDA value		The standard deviation of filtered EDA signal
	Mean tonic EDA level		The mean value of tonic EDA signal
	Max tonic EDA value		The minimum value of tonic EDA signal
	Min tonic EDA value		The maximum value of tonic EDA signal
	Std tonic EDA value		The standard deviation of tonic EDA signal
	Mean amplitude of NS-SCRs		The mean amplitude of NS-SCRs (computed from phasic EDA signal)
	Frequency of NS-SCRs		The number of NS-SCRs per minute (computed from phasic EDA signal)
ECG/ RESP	Mean Rate	Time domain	The mean number of cardiac cycles per minute
	Mean		The mean time of IBIs/BBs
	Median		The median of the absolute values of the successive differences between adjacent IBIs/BBs
	MAD		The mean absolute deviation of IBIs/BBs
	SD		The standard deviation of IBIs/BBs
	SDSD		The standard deviation of the successive differences between adjacent IBIs/BBs
	CV		The Coefficient of Variation, i.e. the ratio of SD divided by Mean
	mCV		Median-based Coefficient of Variation, i.e. the ratio of MAD divided by Median
	RMSSD		The square root of the mean of the sum of successive differences between adjacent IBIs/BBs
	CVSD		The coefficient of variation of successive differences; the RMSSD divided by Mean IBI
	LF	Frequency domain	The spectral power density pertaining to low frequency band (0.04 to.15 Hz)
	HF		The spectral power density pertaining to high frequency band (0.15 to 0.4 Hz)
	LF/HN		The ratio of LF to HF
	SD1	Non-linear domain	Measure of the IBIs/BBs spread on the Poincare plot perpendicular to the line of identity (short-term fluctua- tions)
	SD2		Measure of the IBIs/BBs spread on the Poincare plot along the line of identity (long-term fluctuations)
	SD2/SD1		Ratio between long and short term fluctuations of IBIs (SD2 divided by SD1)
	ApEn		Approximate entropy
ECG	pNN50	Time domain	The proportion of successive IBIs greater than 50 ms, out of the total number of IBIs
	pNN20		The proportion of successive IBIs greater than 20 ms, out of the total number of IBIs
	TINN		The baseline width of IBIs distribution obtained by triangular interpolation
	HTI		The HRV triangular index, measuring the total number of IBIs divided by the height of the IBIs histogram
	IQR		The interquartile range (IQR) of the RR intervals
	SDNNI1(2)		The mean of the standard deviations of RR intervals extracted from 1(2)-minute(s) segments of time series data
	SDANN1(2)		The standard deviation of average RR intervals extracted from 1(2)-minute(s) segments of time series data
	VHF	Frequency domain	Variability, or signal power, in very high frequency (0.4 to 0.5 Hz)
	LFn		The normalised low frequency, obtained by dividing the low frequency power by the total power
	HFn		The normalised high frequency, obtained by dividing the low frequency power by the total power
	LnHF		The log transformed HF
	CSI	Non-linear domain	The Cardiac Sympathetic Index
	CVI		The Cardiac Vagal Index
	CSI modified		The modified CSI obtained by dividing the square of the longitudinal variability by its transverse variability.
	S		Area of ellipse described by SD1 and SD2
	SampEn		Sample entropy
	PIP		Percentage of inflection points of the RR intervals series.
	IALS		Inverse of the average length of the acceleration/deceleration segments
	PSS		Percentage of short segments
	PAS		Percentage of IBIs in alternation segments
	GI		Guzik's Index
	SI		Slope Index
	AI		Area Index
	PI		Porta's Index
	C1d/C1a		Indices of respectively short-term HRV deceleration/acceleration
	SD1d/SD1a		Short-term variance of contributions of decelerations and accelerations
	C2d/C2a		Indices of respectively long-term HRV deceleration/acceleration
	SD2d/SD2a		Long-term variance of contributions of decelerations and accelerations
	Cd/Ca		Total contributions of heart rate decelerations and accelerations to HRV
	SDNNd/SDNNa		Total variance ofcontributions of heart rate decelerations and accelerations to HRV
	DFA alpha1 (2)		The monofractal detrended fluctuation analysis of the HR signal corresponding to short (long)-term correlations
	DFA alpha1 (2) ExpRange		Range of singularity exponents, corresponding to the width of the singularity spectrum from the monofractal detrended fluctuation analysis of the HR signal, corresponding to short(long)-term correlations

(*continued on next page*)

Table 9 (*continued*)

Signal	Indicator	Domain	Description
	DFA alpha1 (2) DimRange		Range of singularity dimensions, corresponding to the height of the singularity spectrum from the monofractal detrended fluctuation analysis of the HR signal, corresponding to short(long)-term correlations
	DFA alpha1 (2) ExpMean		Mean of singularity exponents from the monofractal detrended fluctuation analysis of the HR signal, corresponding to short(long)-term correlations
	DFA alpha1 (2) DimMean		Mean of singularity dimension from the monofractal detrended fluctuation analysis of the HR signal, corresponding to short(long)-term correlations
	ShanEn		Shannon entropy
	FuzzyEn		Fuzzy entropy
	MSE		Multiscale entropy
	CMSE		Composite multiscale entropy
	RCMSE		Refined composite multiscale entropy
	CD		Correlation dimension
	HFD		Higuchi's Fractal Dimension of the HR signal
	KFD		The Katz's Fractal Dimension of the HR signal
	LZC		The Lempel-Ziv complexity of the HR signal
RESP	Mean amplitude	Time domain	The mean respiratory amplitude.
	Phase Duration Inspiration		The average inspiratory duration
	Phase Duration Expiration		The average expiratory duration
	Phase Duration Ratio		The inspiratory-to-expiratory time ratio (I/E)
RSA	Mean (P2T)		Mean of RSA estimates (peak-to-trough method)
	Mean Log (P2T)		The logarithm of the mean of RSA estimates (peak-to-trough method)
	SD (P2T)		The standard deviation of all RSA estimates (peak-to-trough method) 31
	Mean (Gates)		Mean of RSA estimates (Gates method)
	Mean Log (Gates)		The logarithm of the mean of RSA estimates (Gates method)
	SD (Gates)		The standard deviation of all RSA estimates (Gates method)
	PorgesBohrer		The Porges-Bohrer estimate of RSA, optimal when the signal to noise ratio is low, in ln(ms2)

environment, which had no traffic and a slightly changing landscape throughout the scenario (Thiffault and Bergeron, 2003b; Jarosch et al., 2019). In contrast, the urban scenario may have been more stimulating because of other road users (cars and pedestrians), a regularly changing trajectory, and different buildings throughout the drive.

Interestingly, participants who drove first in the urban environment reported being sleepier in the rural environment later. The higher complexity of the urban environment might have increased the feeling of sleepiness in the second part of the experiment. Driving in the urban area first might be more critical for drivers' state afterwards. Attention should be paid to the scenario order in further research.

Affect

In link with results obtained on sleepiness, participants reported being less aroused after 30 min of conditionally automated driving, specifically in the monotonous rural scenario. Besides, no effect of sleep deprivation or driving environment was found on subjective measures of arousal. This was further investigated through statistical analysis and classification tasks with objective physiological measures.

With regard to drivers' valence, the statistical analysis revealed that drivers were more negative after driving than before. This might also be explained by the monotonous nature of the experiment. Indeed, it might have been perceived as boring by participants, who might have expected to take over more often and perform more activities during the ride. ***To summarize, (H2b) is validated, whereas (H1b) and (H3b) are refuted.***

Drivers' physiological state

Inter-subjects variability

The statistical analysis revealed a significant effect of baseline as covariate for the selected physiological measures. This means that there was high variability in the physiological state among drivers at rest, which may be due to extraneous individual differences (Cacioppo et al., 2007). This suggests that it is relevant to take into account the physiological state at rest to evaluate accurately the driver's state while driving. This has been done through feature engineering in the classification pipeline implemented in this study (explained in section 3.7.2).

Sleep deprivation

Statistical analysis revealed no significant effect of sleep deprivation on the selected physiological measures, though a significant effect was found on self-reported measures. ***It means that (H1c) is refuted.*** Whereas previous research showed that the selected measures are appropriate for measuring changes in physiological state related to sleep deprivation (Posada-Quintero et al., 2017; Liu et al., 2015; Burton et al., 2010; Chua et al., 2012; Schmitt et al., 2015), sleep deprivation might have been too mild to see an effect on the drivers' physiological state. Also, indicators were calculated from signals corresponding to the entire duration of a scenario (30 min). The effect of sleep deprivation might have faded over time and so the values measured between the experimental groups were close.

This is confirmed by the low accuracy achieved by ML models with a segmentation level of 1. It revealed that sleep deprivation could be predicted with only 71 %-accuracy from features calculated over the entire driving period. Sensor fusion did not significantly increase the performance of the model. However, signal segmentation increased the model accuracy. Indeed, sleep deprivation could be predicted with an accuracy of 99 % using a segmentation level of 30 and the EDA signal as the unique input of a neural network. This means it was easier to predict sleep deprivation using EDA features calculated from 30 windows of 1-minute, rather than one large 30-minute window. Segmentation provides the model with a greater amount of information (more training samples), but also with a greater quality of information in having more details on the evolution of the driver's state over time. However, previous research showed that the smallest window does not necessarily yield the best model performance (Meteier et al., 2021a). The score achieved in this study can serve as the baseline for further research as no previous studies predicted sleep-related fatigue alone from physiological signals using ML.

The post-hoc analysis on feature importance revealed that the frequency and amplitude of non-specific skin conductance responses were found among the ten most predictive features (see column CT1 in Table 8). This is consistent with the findings of Rault et al. (2019) who found that phasic EDA features are relevant indicators of sleep deprivation.

Driving environment and scenario order

Regarding the tonic EDA level as a measure of drivers' arousal, results showed that it was consistently higher in the second scenario, regardless of which scenario participants drove first (see Fig. 5). Even when driving in the rural environment after the urban one, the arousal kept increasing, though one could have thought that the arousal would have decreased since the rural scenario after can be considered as monotonous. Drivers did not perform actively the driving task but the sole task of monitoring the environment is still arousing. The increase of arousal in the second scenario might also be explained by the takeover request received at the end of the first scenario. It might have induced an increase in arousal, which lasted later on during the second scenario. This should be considered for further research in driver's state evaluation, especially for long periods of automated driving.

Regarding SDNN and RSA as measures of parasympathetic activity, the significant effect of driving environment and scenario order also showed a greater parasympathetic dominance in the second part of the experiment, regardless of which scenario was presented first. This might correspond to an increased fatigue experienced by drivers after one hour of monotonous automated driving. ***As the driving environment alone did not influence significantly the drivers' physiological state, (H3c) is refuted.***

The classification task with ML techniques confirms that no significant effect of the driving environment was found in the statistical analysis. Indeed, only 55 %-accuracy could be achieved by the model with a segmentation level of 1, using the three signals. Sensor fusion slightly improved the performance of the model. For this classification task, segmentation also improved model performance, as 85 %-accuracy was achieved with the only EDA signal and a segmentation level of 30. The difference detected by the model between the two conditions may come from the complexity of the urban environment and its dynamic character (traffic with more changing factors), compared to the monotonous aspect of the rural environment. The urban environment could also induce an increase in mental load (Lyu et al., 2017) and/or stress (Healey and Picard, 2005), which may be reflected in the physiological features.

The accuracy achieved in this study is lower than the one obtained by Healey and Picard (2005) or Chen et al. (2017), ***as assumed in (H3d)***. They used data collected in real-driving conditions to train ML models, which might increase driving stress. Indeed, a danger or accident cannot really happen in the simulation, which might play a role in physiological activation. Besides, results are close to those obtained by Darzi et al. (2018) who predicted with 86.8 %-accuracy the driving environment (highway vs. town) only from physiological signals in a driving simulator. Unlike these previous studies in which drivers had to drive manually, the car drove in conditional automation in this experiment. Even though drivers were not actively driving and only had to observe the environment, results show that similar performance can be achieved to evaluate the physiological change induced by the environment at L3-SAE, compared to manual driving.

Sleepiness

The prediction of sleepiness proved to be the most complicated task to accomplish for ML algorithms. The best classifier was able to predict sleepiness towards the end of the monotonous rural scenario with 73 %-accuracy. The best model used EDA features calculated from the first and last five 1-minute segments of the drive (segmentation level of 30). The physiological state of the drivers may have changed a little after 30 min of automated driving, but not enough to bring the model close to 100 %-accuracy. It suggests that some drivers would still be alert enough and potentially ready to take over control, despite low engagement during a 30-minute period of conditionally automated driving. It should be tested whether the model performance would increase as car travel time increases. In fact, a recent study showed that 50 min of monotonous driving could already affect drivers' takeover quality (Jarosch et al., 2019). The accuracy achieved is far from the one achieved in previous L2-SAE studies (Kundinger and Riener, 2020; Kundinger et al., 2020), ***which refutes (H2c)***. The duration of the scenarios (30 vs. 45 min), the experiment hour (at times of low alertness in previous studies), and other limitations (see below) can explain this difference. Further research should be conducted to confirm whether a decrease in alertness can be detected more accurately with physiological signals and ML techniques at L3-SAE driving.

Limitations and further work

This study includes several limitations which are discussed here. Even though the statistical analysis revealed that sleep-deprived participants felt more drowsy before driving, some of them might not have respected instructions to reduce their sleep. This could be addressed by using a sleep tracker to get an objective measure of sleeping time and make sure participants respected the instructions.

In this study, a driving simulation was used since it is the safest way to experimentally manipulate driver fatigue without endangering the driver while providing environments very similar to real-world driving scenarios. However, results obtained in this study should be replicated by other studies in real driving situations at L3-SAE. Besides, the ML model predicting the driving environment might have been affected by problems encountered in the urban environment. Some participants had to take over control without the car requesting to do so, because of a collision with another car or a wrong autopilot trajectory. It may have contributed to keeping the affected drivers more alert in this scenario and induced more noise in the collected physiological data. In addition, some participants were surrounded by machine noises, which may have also affected their physiological state and slightly skewed the results.

As further work, other machine or deep learning algorithms, especially those efficient on time series data, could be tested to increase the accuracy obtained in this study or perform a finer-grained analysis to find factors related to safety issues (Mirzahossein et al., 2022). Regression techniques could also be applied to physiological data to obtain a finer evaluation of the drivers' sleepiness level using KSS as ground truth. Predictions of ML models implemented in this study could be used to alert the driver when a large change in physiological arousal is detected. Based on that, an intelligent system could send a salient warning if the driver's physiological state is critical. Also, the human-vehicle interaction could be adapted implicitly by sending biofeedback adapted to the drivers' state (calming or stimulating music, cold or warm light, etc..) (Capallera et al., 2021; Daher et al., 2021; de Salis et al., 2020).

Conclusion

To understand how fatigue and driving environment play a role in the driver's physiological state specifically in conditionally automated driving (L3-SAE), a driving simulator study was conducted with 63 drivers. The vehicle drove in two different environments (rural vs. urban) for 30 min each, with the rural environment inducing sleepiness. Drivers had to take over control when requested by the car, which occurred at the end of each scenario. Half of the subjects were mildly sleep-deprived the night before the experiment and slept less than six hours.

Participants reported being subjectively sleepier and less aroused after 30 min of automated driving in the monotonous scenario, and sleep-deprived drivers reported to be sleepier before and after the drive. Regarding drivers' physiological state, the statistical analysis did not reveal a significant effect of sleep deprivation and driving environment on selected physiological measures. However, classification tasks with ML techniques showed that mild sleep deprivation, driving environment (rural vs. urban), and sleepiness could be respectively predicted with 99 %, 85 %, and 73 %-accuracy. The low accuracy achieved for sleepiness prediction suggests that not all drivers are drowsy after only 30 min of conditionally automated driving. EDA features calculated from multiple 1-minute segments yielded the best results. Segmentation increased the accuracy of ML classifiers. Some of the most relevant physiological features for predicting the manipulated factors are also reported in this work. In particular, the majority of the most predictive features of fatigue (sleep deprivation and sleepiness) were corrected with baseline, which might be useful for further research.

The results highlight that certain factors such as the type of environment, drowsiness or sleep deprivation, however mild, can be detected using ML techniques and physiological data. Robust and continuous driver assessment is necessary to reduce the number of accidents on the roads. European regulations align with this since new vehicles sold on the market now must be equipped with driver drowsiness and attention warning systems. The latter often use facial analysis (eyelid opening, for example) and steering wheel movement as input data, but they are not relevant for measuring the driver's state at levels 3 or 4 of automation as the driver does not perform the driving task anymore. This empirical study therefore encourages car manufacturers to incorporate the measurement of physiological indicators into the design of smart driver assessment systems, to warn drivers in good time before critical situations arise.

CRediT authorship contribution statement

Quentin Meteier: Conceptualization, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Reńee Favre:** Data curation, Methodology, Writing – review & editing. **Sofia Viola:** Data curation, Methodology, Writing – review & editing. **Marine Capallera:** Software, Writing – review & editing. **Leonardo Angelini:** Supervision, Writing – review & editing. **Elena Mugellini:** Funding acquisition, Project administration, Writing – review & editing. **Andreas Sonderegger:** Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data collected in this study are available here: https://www.sciencedirect.com/science/article/pii/S2352340923001452

Acknowledgments

This work has been supported by the Hasler Foundation in the framework of the AdVitam project. The authors would like to thank Yagmur Baran for her help in the data collection, as well as Emmanuel de Salis, Stefano Carrino, and Marino Widmer for their implication in this project.

References

Ackermann, S., Hartmann, F., Papassotiropoulos, A., de Quervain, D.J.F., Rasch, B., 2015. No Associations between Interindividual Differences in Sleep Parameters and Episodic Memory Consolidation. Sleep 38, 951–959. https://doi.org/10.5665/sleep.4748.

Ahlström, C., Zemblys, R., Jansson, H., Forsberg, C., Karlsson, J., Anund, A., 2021. Effects of partially automated driving on the development of driver sleepiness. Accident; analysis and prevention 153, 106058.

Akerstedt, T., Gillberg, M., 1990. Subjective and objective sleepiness in the active individual. Int. J. Neurosci. 52 (1–2), 29–37.

Altemus, M., Rao, B.N., Dhabhar, F.S., Ding, W., Granstein, R.D., 2001. Stress-induced changes in skin barrier function in healthy women. J. Invest. Dermatol. 117 (2), 309–317.

Awais, M., Badruddin, N., Drieberg, M., 2017. A hybrid approach to detect driver drowsiness utilizing physiological signals to improve system performance and wearability 17. doi:10.3390/s17091991.

Bestelmeyer, P.E.G., Kotz, S.A., Belin, P., 2017. Effects of emotional valence and arousal on the voice perception network. URL Soc. Cognit. Affect. Neurosci. 12, 1351–1358.

Borisenkov, M.F., Perminova, E.V., Kosova, A.L., 2010. Chronotype, sleep length, and school achievement of 11- to 23-year-old students in northern european russia. Chronobiol. Int. 27, 1259–1270.

Bundele, M.M., Banerjee, R., 2009. An SVM classifier for fatigue detection using skin conductance for use in the BITS-lifeguard wearable computing system, in: 2009 Second International Conference on Emerging Trends in Engineering & Technology, IEEE. pp. 934–939. URL: http://ieeexplore.ieee.org/document/5395425/, doi: 10.1109/ICETET.2009.29.

Burton, A.R., Rahman, K., Kadota, Y., Lloyd, A.R., Vollmer-Conna, U., 2010. Reduced heart rate variability predicts poor sleep quality in a case–control study of chronic fatigue syndrome. Experimental Brain Research 204, 71–78.

Cacioppo, J.T., Tassinary, L.G., Berntson, G.G., 2007. Handbook of psychophysiology.

Cai, A.W.T., Manousakis, J.E., Singh, B., Kuo, J., Jeppe, K.J., Francis-Pester, E., Shiferaw, B., Beatty, C., Rajaratnam, S.M., Lenńe, M.G., Howard, M.E., Anderson, C., 2021. On-road driving impairment following sleep deprivation differs according to age. Sci. Rep. 11.

Capallera, M., Meteier, Q., Koch, K., Funk, M., Kamali, M.E., Daher, K., Khaled, O.A., Mugellini, E., 2021. Workshop for designing biofeedback of driver's state and emotion in automated vehicles. 13th International Conference on Automotive User Interfaces and Interactive Vehicular Applications.

Chen, L., Zhao, Y., fei Ye, P., Zhang, J., Zou, J., 2017. Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. Expert Syst. Appl. 85, 279–291.

Chua, E.C.P., Tan, W., Yeo, S.C., Lau, P., Lee, I.T.G., Mien, I.H., Puvanendran, K., Gooley, J.J., 2012. Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation. Sleep 35 (3), 325–334.

Costa, G., 1996. The impact of shift and night work on health. Appl. Ergon. 27 (1), 9–16.

Daher, K., Capallera, M., Lucifora, C., Casas, J., Meteier, Q., Kamali, M.E., Ali, A.E., Grosso, G.M., Chollet, G., Khaled, O.A., Mugellini, E., 2021. Empathic interactions in automated vehicles #empathicchi. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.

Darzi, A., Gaweesh, S.M., Ahmed, M.M., Novak, D., 2018. Identifying the causes of drivers' hazardous states using driver characteristics, vehicle kinematics, and physiological measurements 12, 568. URL :https://www.frontiersin.org/article/10.3389/fnins.2018.00568/full, doi:10.3389/fnins.2018.00568.

de Naurois, C.J., Bourdin, C., Stratulat, A., Diaz, E., Vercher, J.L., 2017. Detection and prediction of driver drowsiness using artificial neural network models. URL Accid. Anal. Prev. 126, 95–104.

de Salis, E., Capallera, M., Meteier, Q., Angelini, L., Khaled, O.A., Mugellini, E., Widmer, M., Carrino, S., 2020. Designing an ai-companion to support the driver in highly autonomous cars, in: HCI.

di Flumeri, G., Borghini, G., Aric'o, P., Sciaraffa, N., Lanzi, P., Pozzi, S., Vignali, V., Lantieri, C., Bichicchi, A., Simone, A., Babiloni, F., 2018. Eeg-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. Front. Human Neurosci. 12.

Du, N., Zhou, F., Pulver, E., Tilbury, D.M., Robert, L.P., Pradhan, A.K., Yang, J.X., 2020. Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. PsychRN: Other Cognitive Psychology (Topic).

Fujiwara, K., Abe, E., Kamata, K., Nakayama, C., Suzuki, Y., Yamakawa, T., Hiraoka, T., Kano, M., Sumi, Y., Masuda, F., Matsuo, M., Kadotani, H., 2019. Heart rate variability-based driver drowsiness detection and its validation with EEG 66, 1769–1778. doi:10.1109/TBME.2018.2879346. conference Name: IEEE Trans. Biomed. Eng.

Garde, A.H., Begtrup, L.M., Bjorvatn, B., Bonde, J.P.E., Hansen, J., Hansen, Å.M., Härmä, M., Jensen, M.A., Kecklund, G., Kolstad, H.A., Larsen, A.D., Lie, J.A., Moreno, C.R.C., Nabe-Nielsen, K., Sallinen, M., 2020. How to schedule night shift work in order to reduce health and safety risks. Scandinavian J. Work, Environ. Health 46, 557 – 569.

Gates, K.M., Gatzke-Kopp, L.M., Sandsten, M., Blandon, A.Y., 2015. Estimating time-varying rsa to examine psychophysiological linkage of marital dyads. Psychophysiology 52, 1059–1065. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/psyp.12428, doi:https://doi.org/10.1111/psyp.12428, arXiv:https://onlinelibrary.wiley.com/doi/pdf/.

Feldhütter, A., Hecht, T., Kalb, L., Bengler, K., 2018. Effect of prolonged periods of conditionally automated driving on the development of fatigue: with and without non-driving-related activities. Cogn. Tech. Work 21, 33–40.

Gilbert, S.P., Weaver, C.C., 2010. Sleep quality and academic performance in university students: A wake-up call for college psychologists. J. Coll. Stud. Psychother. 24, 295–306.

Guo, H., Zhang, Y., Cai, S., Chen, X., 2021. Effects of level 3 automated vehicle drivers' fatigue on their take-over behaviour: A literature review. J. Adv. Transportat. 2021.

Healey, J., Picard, R.W., 2005. Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems 6, 156–166.

Heikoop, D., Hagenzieker, M., Mecacci, G., Calvert, S., Santoni de Sio, F., Arem, B., 2019. Human behaviour with automated driving systems: a quantitative framework for meaningful human control. Theor. Issues Ergon. Sci. 20, 1–21. https://doi.org/10.1080/1463922X.2019.1574931.

Hertach, P., Uhr, A., Niemann, S., Huwiter, K., Achermann-Stürmer, Y., 2020. Impaired driving ability of motor vehicle drivers. BFU, Swiss Council for Accident Prevention, Bern.

Jarosch, O., Bellem, H., Bengler, K., 2019. Effects of task-induced fatigue in prolonged conditional automated driving. Human Factors: the Journal of Human Factors and Ergonomics Society 61, 1186–1199.

Jewett, M.E., Dijk, D.J., Kronauer, R.E., Dinges, D.F., 1999. Dose-response relationship between sleep duration and human psychomotor vigilance and subjective alertness. Sleep 22 (2), 171–179.
Kaida, K., Akerstedt, T., Kecklund, G., Nilsson, J.P., Axelsson, J., 2007. Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task. Ind. Health 45 (4), 520–526.
Katona, P.G., Jih, F., 1975. Respiratory sinus arrhythmia: noninvasive measure of parasympathetic cardiac control. J. Appl. Physiol. 39 (5), 801–805.
Kiashari, S.E.H., Nahvi, A., Bakhoda, H., Homayounfard, A., Tashakori, M., 2020. Evaluation of driver drowsiness using respiration analysis by thermal imaging on a driving simulator 79, 17793–17815. URL: http://link.springer.com/10.1007/s11042-020-08696-x, doi:10.1007/s11042-020-08696-x.
Kundinger, T., Riener, A., 2020. The potential of wrist-worn wearables for driver drowsiness detection: A feasibility analysis, in: Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, ACM. pp. 117–125 URL: https://dl.acm.org/doi/10.1145/3340631.3394852, doi:10.1145/3340631.3394852.
Kundinger, T., Sofra, N., Riener, A., 2020. Assessment of the potential of wrist-worn wearable sensors for driver drowsiness detection 20, 1029. URL: https://www.mdpi.com/1424-8220/20/4/1029, doi:10.3390/s20041029.
Larue, G.S., Rakotonirainy, A., Pettitt, A.N., 2011. Driving performance impairments due to hypovigilance on monotonous roads. Accid. Anal. Prev. 43 (6), 2037–2046.
Lee, H., Lee, J., Shin, M., 2019. Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots , 15.
Li, G., Chung, W.Y., 2013. Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier 13, 16494–16511. URL: http://www.mdpi.com/1424-8220/13/12/16494, doi:10.3390/s131216494.
Liu, J.C.J., Verhulst, S., Massar, S.A.A., Chee, M.W.L., 2015. Sleep deprived and sweating it out: the effects of total sleep deprivation on skin conductance reactivity to psychosocial stress. Sleep 38 (1), 155–159.
Llaneras, R.E., Salinger, J.A., Green, C.A., 2017. Human factors issues associated with limited ability autonomous driving systems: Drivers' allocation of visual attention to the forward roadway.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions, in: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 30. Curran Associates, Inc., pp. 4765–4774.
Lyu, N., Xie, L., Wu, C., Fu, Q., Deng, C., 2017. Driver's cognitive workload and driving performance under traffic sign information exposure in complex environments: A case study of the highways in china. Int. J. Environ. Res. Public Health 14.
Makowski, D., Pham, T., Lau, Z.J., Brammer, J.C., Lespinasse, F., Pham, H., Schölzel, C., Chen, S.H.A., 2021. Neurokit2: A python toolbox for neurophysiological signal processing. Behavior research methods.
May, J.F., Baldwin, C.L., 2009. Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies. Transportation Research Part F: Traffic Psychology and Behaviour 12, 218–224. URL:http://www.sciencedirect.com/science/article/pii/S1369847808001009, doi: 10.1016/j.trf.2008.11.005.
Meteier, Q., Capallera, M., Ruffieux, S., Angelini, L., Khaled, O.A., Mugellini, E., Widmer, M., Sonderegger, A., 2021a. Classification of drivers' workload using physiological signals in conditional automation. Front. Psychol. 12 https://doi.org/10.3389/fpsyg.2021.596038.
Meteier, Q., de Salis, E., Capallera, M., Widmer, M., Angelini, L., Khaled, O.A., Sonderegger, A., Mugellini, E., 2021b. Relevant physiological indicators for assessing workload in conditionally automated driving, through three-class classification and regression. Front. Comput. Sci. 3 https://doi.org/10.3389/fcomp.2021.775282.
Meteier, Q., Kindt, M., Angelini, L., Khaled, O.A., Mugellini, E., 2022. Non-intrusive contact respiratory sensor for vehicles. Sensors (Basel, Switzerland) 22.
Mirzahossein, H., Sashurpour, M., Hosseinian, S.M., Gilani, V.N.M., 2022. Presentation of machine learning methods to determine the most important factors affecting road traffic accidents on rural roads. Front. Struct. Civ. Eng. 16, 657–666.
Moller, H.J., Kayumov, L., Shapiro, C., 2003. Microsleep episodes, attention lapses and circadian variation in psychomotor performance in a driving simulation paradigm.
Mosteller, F., Tukey, J.W., 1968. Data analysis, including statistics, in: Lindzey, G., Aronson, E. (Eds.), Handbook of Social Psychology, Vol. 2, Addison-Wesley.
Nakajima, Y., Tanaka, K., 2017. Effects of active and passive secondary tasks in a take-over situation during automated driving. In: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1161–1166. https://doi.org/10.1109/SMC.2017.8122769.
Ness, T.E.B., Saksvik-Lehouillier, I., 2018. The relationships between life satisfaction and sleep quality, sleep duration and variability of sleep in university students.
Swiss Federal Statistical Office, S.F.S.O., 2020. Road traffic accidents.
Okano, K., Kaczmarzyk, J.R., Dave, N., Gabrieli, J.D.E., Grossman, J.C., 2019. Sleep quality, duration, and consistency are associated with better academic performance in college students. NPJ Science of Learning 4.
Palmer, C.A., Alfano, C.A., 2017. Sleep and emotion regulation: An organizing, integrative review. Sleep Med. Rev. 31, 6–16.
Patel, M., Lal, S., Kavanagh, D., Rossiter, P., 2011. Applying neural network analysis on heart rate variability data to assess driver fatigue 38, 7235–7242. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957417410013916, doi:10.1016/j.eswa.2010.12.028.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Philip, P., Sagaspe, P., Taillard, J., Moore, N., Guilleminault, C., Sanchez-Ortuno, M., Akerstedt, T., Bioulac, B.H., 2003. Fatigue, sleep restriction, and performance in automobile drivers: a controlled study in a natural environment. Sleep 26 (3), 277–280.
Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Akerstedt, T., Charles, A., Bioulac, B.H., 2005. Fatigue, sleepiness, and performance in simulated versus real driving conditions. Sleep 28 (12), 1511–1516.
Pires, G.N., Bezerra, A.G., Tufik, S., Andersen, M.L., 2016. Effects of acute sleep deprivation on state anxiety levels: a systematic review and meta-analysis. Sleep Med. 24, 109–118.
Posada-Quintero, H.F., Bolkhovsky, J.B., Reljin, N., Chon, K.H., 2017. Sleep deprivation in young and healthy subjects is more sensitively identified by higher frequencies of electrodermal activity than by skin conductance level evaluated in the time domain. Front. Physiol. 8.
Potter, G.D.M., Skene, D.J., Arendt, J., Cade, J.E., Grant, P.J., Hardie, L.J., 2016. Circadian rhythm and sleep disruption: Causes, metabolic consequences, and countermeasures. Endocr. Rev. 37, 584–608.
Radlmayr, J., Feldhütter, A., Frey, A., Jarosch, O., Marberger, C., Naujoks, F., Wehlack, V., Bengler, K., 2018. Drowsiness and fatigue in conditionally automated driving - towards an integrative framework.
Rault, C., Sangaré, A., Diaz, V., Ragot, S., Frat, J.P., Raux, M., Similowski, T., Robert, R., Thille, A.W., Drouot, X., 2019. Impact of sleep deprivation on respiratory motor output and endurance: A physiological study. American journal of respiratory and critical care medicine.
Reimer, B., D'Ambrosio, L.A., Coughlin, J.F., 2007. Secondary analysis of time of day on simulated driving performance. J. Saf. Res. 38 (5), 563–570.
Rigas, G., Goletsis, Y., Bougia, P., Fotiadis, D.I., 2011. Towards driver's state recognition on real driving conditions 2011, e617210. URL: https://www.hindawi.com/journals/ijvt/2011/617210/, doi:10.1155/2011/617210. publisher: Hindawi.
Rowden, P.J., Matthews, G., Watson, B., Biggs, H.C., 2011. The relative impact of work-related stress, life stress and driving environment stress on driving outcomes. Accid. Anal. Prev. 43 (4), 1332–1340.
Russell, J.A., 1980. A circumplex model of affect. URL J. Personal. Social Psychol. 39, 1161–1178.
SAE, 2018. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles.
Schömig, N., Hargutt, V., Neukum, A., Petermann-Stock, I., Othersen, I., 2015. The Interaction Between Highly Automated Driving and the Development of Drowsiness. Procedia Manufacturing 3, 6652–6659. https://doi.org/10.1016/j.promfg.2015.11.005.
Scharfe, M.S.L., Zeeb, K., Rußwinkel, N., 2020. The impact of situational complexity and familiarity on takeover quality in uncritical highly automated driving scenarios. Inf 11, 115.
Schmitt, L., Regnard, J., Millet, G.P., 2015. Monitoring fatigue status with hrv measures in elite athletes: An avenue beyond rmssd? Frontiers in Physiology 6.
Sharma, M.K., Bundele, M.M., 2015. Design & analysis of performance of k-means algorithm for cognitive fatigue detection in vehicular drivers using skin conductance signal. In: In: 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), pp. 707–712.
Shen, J., Barbera, J., Shapiro, C.M., 2006. Distinguishing sleepiness and fatigue: focus on definition and measurement.Sleep Medicine Reviews 10, 63–76. URL: https://www.sciencedirect.com/science/article/pii/S1087079205000444, Doi: 10.1016/j.smrv.2005.05.004.
Shinar, D., Compton, R., 2004. Aggressive driving: an observational study of driver, vehicle, and situational variables. Accid. Anal. Prev. 36 (3), 429–437.
Shinar, D., 1978. Psychology on the road: The human factor in traffic safety.
Sonderegger, A., Heyden, K., Chavaillaz, A., Sauer, J., 2016. Anisam & aniavatar: Animated visualizations of affective states, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 4828–4837. URL: https://doi.org/10.1145/2858036.2858365, doi:10.1145/2858036.2858365.
Thiffault, P., Bergeron, J., 2003a. Fatigue and individual differences in monotonous simulated driving. Pers. Individ. Differ. 34, 159–176.
Thiffault, P., Bergeron, J., 2003b. Monotony of road environment and driver fatigue: a simulator study. Accid. Anal. Prev. 35 (3), 381–391.
Tomaso, C.C., Johnson, A.B., Nelson, T.D., 2020. The effect of sleep deprivation and restriction on mood, emotion, and emotion regulation: Three meta-analyses in one. Sleep.
Underwood, G.M., 2007. Visual attention and the transition from novice to advanced driver. Ergonomics 50, 1235–1249.
Valdez, P., 2019. Circadian Rhythms in Attention. The Yale Journal of Biology and Medicine 92, 81–92. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6430172/.
van Hooft, E.A.J., van Hooff, M.L.M., 2018. The state of boredom: Frustrating or depressing? Motiv. Emot. 42, 931–946.
Vicente, J., Laguna, P., Bartra, A., Bailón, R., 2011. Detection of driver's drowsiness by means of HRV analysis, in: 2011 Computing in Cardiology, pp. 89–92.
Vogelpohl, T., Kühn, M., Hummel, T., Vollrath, M., 2019. Asleep at the automated wheel-sleepiness and fatigue during highly automated driving. Accident; Analysis and Prevention 126, 70–84.
Wandtner, B., Schömig, N., Schmidt, G., 2018. Effects of Non-Driving Related Task Modalities on Takeover Performance in Highly Automated Driving. Human Factors 60, 870–881. URL: https://doi.org/10.1177/0018720818768199. Publisher: SAGE Publications Inc.
Wang, D., Shen, P., Wang, T., Xiao, Z., 2017. Fatigue detection of vehicular driver through skin conductance, pulse oximetry and respiration: A random forest

classifier, in: 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN), pp. 1162–1166. doi:10.1109/ICCSN.2017.8230293. ISSN: 2472-8489.

Zepf, S., Dittrich, M., Herńandez, J., Schmitt, A., 2019. Towards empathetic car interfaces: Emotional triggers while driving. Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.

Zhou, F., Alsaid, A., Blommer, M., Curry, R., Swaminathan, R., Kochhar, D.S., Talamonti, W.J., Tijerina, L., Lei, B., 2020. Driver fatigue transition prediction in highly automated driving using physiological features. Expert Syst. Appl. 147, 113204.

Zhou, F., Alsaid, A., Blommer, M., Curry, R., Swaminathan, R., Kochhar, D.S., Talamonti, W.J., Tijerina, L., 2021. Predicting driver fatigue in monotonous automated driving with explanation using gpboost and shap. Int. J. Human-Comput. Interact. 38, 719–729.