Detecting Driver Sleepiness Using Consumer Wearable Devices in Manual and Partial Automated Real-Road Driving

Citation for the original published paper (version of record):
Lu, K., Karlsson, J., Sjörs, A. et al (2022). Detecting Driver Sleepiness Using Consumer Wearable Devices in Manual and Partial Automated
Real-Road Driving. IEEE Transactions on Intelligent Transportation Systems, 23(5): 4801 -4810.
http://dx.doi.org/10.1109/TITS.2021.3127944

Detecting Driver Sleepiness Using Consumer Wearable Devices in Manual and Partial Automated Real-Road Driving

Ke Lu, *Member, IEEE*, Johan Karlsson, Anna Sjörs Dahlman, Bengt Arne Sjöqvist, and Stefan Candefjord

***Abstract*— Driver sleepiness constitutes a well-known traffic safety risk. With the introduction of automated driving systems, the chance of getting sleepy and even falling asleep at wheel could increase further. Conventional sleepiness detection methods based on driving performance and behavior may not be available under automated driving. Methods based on physiological measurements such as heart rate variability (HRV) becomes a potential solution under automated driving. However, with reduced task load, HRV could potentially be affected by automated driving. Therefore, it is essential to investigate the influence of automated driving on the relation between HRV and sleepiness. Data from real-road driving experiments with 43 participants were used in this study. Each driver finished four trials with manual and partial automated driving under normal and sleep-deprived condition. Heart rate was monitored by consumer wearable chest bands. Subjective sleepiness based on Karolinska sleepiness scale was reported at five min interval as ground truth. Reduced heart rate and increased overall variability were found in association with severe sleepy episodes. A binary classifier based on the AdaBoost method was developed to classify alert and sleepy episodes. The results indicate that partial automated driving has small impact on the relationship between HRV and sleepiness. The classifier using HRV features reached area under curve (AUC) = 0.76 and it was improved to AUC = 0.88 when adding driving time and day/night information. The results show that commercial wearable heart rate monitor has the potential to become a useful tool to assess driver sleepiness under manual and partial automated driving.**

***Index Terms*— Heart rate variability, driver sleepiness, real-road driving, automated driving, driver monitoring system, wearable sensors, machine learning.**

Manuscript received November 1, 2020; revised March 17, 2021 and August 18, 2021; accepted September 23, 2021. Date of publication December 10, 2021; date of current version May 3, 2022. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme Projects Mediator under Grant 814735, in part by ADAS&ME under Grant 688900, in part by the Strategic Vehicle Research and Innovation (FFI) Program at the Sweden's Innovation Agency (VINNOVA) under Grant 2020-05157, and in part by the Autoliv Development AB and SmartEye AB. The Associate Editor for this article was C. Ahlstrom. *(Corresponding author: Ke Lu.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Swedish Ethical Review Authority under Application No. Dnr 2019-04813.

Ke Lu, Bengt Arne Sjöqvist, and Stefan Candefjord are with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden, and also with the SAFER Vehicle and Traffic Safety Centre, Chalmers University of Technology, 41756 Gothenburg, Sweden (e-mail: ke.lu@chalmers.se; bengt.arne.sjoqvist@chalmers.se; stefan.candefjord@chalmers.se).

Johan Karlsson is with Autoliv Research, Autoliv Development AB, 44737 Vårgårda, Sweden, and also with the SAFER Vehicle and Traffic Safety Centre, Chalmers University of Technology, 41756 Gothenburg, Sweden (e-mail: Johan.G.Karlsson@autoliv.com).

Anna Sjörs Dahlman is with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden, also with the Swedish National Road and Transport Research Institute, 58330 Linköping, Sweden, and also with the SAFER Vehicle and Traffic Safety Centre, Chalmers University of Technology, 41756 Gothenburg, Sweden (e-mail: anna.dahlman@vti.se).

Digital Object Identifier 10.1109/TITS.2021.3127944

I. Introduction

DRIVER sleepiness and fatigue is a contributing factor to 10–30% of fatal crashes [1]–[3]. Therefore, driver sleepiness monitoring systems have the potential to save many lives. According to Euro NCAP 2025 roadmap [4], driver state monitoring will become one part of their safety assessment. The sleepiness detection systems are typically based on driving performance, such as steering and speed; face analysis, including eye closure, head pose, and eye gaze; and physiological measurements, such as electroencephalography (EEG), electrocardiography (ECG) and electromyography (EMG). Most of the current commercially available sleepiness detection systems are based on driving performance and facial features detection [5].

Applying vehicle automation is another motion towards improving road safety. By having automated longitudinal and lateral control, the driving system relieves the driver from parts of their driving tasks and provides emergency breaking and turning. SAE International defined six levels for driving automation from no automation to full automation [6]. At level 2 (partial driving automation), the driver is responsible for monitoring the environment. At level 3 (conditional driving automation), the driver should be prepared to take over driving when required. Before highly or fully automated (level 4 and 5) vehicles become commercially available, drivers should maintain a state where they are able to take back control of the vehicle at any time. However, fatigue due to underload will probably be more frequent in automated driving when drivers do not have active task engagement [7]–[9]. Therefore, driver fatigue becomes a major concern for partial and conditional automated driving. When steering and speed are partially controlled by the vehicle, the vehicle-based driving performance metrics will not be available for analyzing driver state in partial automation [10]. For higher levels of automation when the driver is not responsible for monitoring the environment, gaze, eyelid closure or head positioning may be less useful as indicators of driver fatigue [11]. These changes challenge current commercially available driver monitoring systems. Adding measurements of physiological indicators of fatigue could be a valuable complement for driver monitoring also in automated driving.

Heart rate variability (HRV) is known to reflect parasympathetic and sympathetic activity. Several sleep laboratory studies show that HRV can be a good indicator for vigilance state under total sleep deprivation [12] and partial sleep deprivation [13], [14]. With the development of unobtrusive sensing techniques, heart rate (HR) and HRV could be measured through wearable sensors [15], [16] or vehicle integrated sensors [17], [18] in a daily driving scenario. Several studies have approached the relationship between driver sleepiness and HRV parameters by building sleepiness classifiers based on HRV features [19]–[26]. High sensitivity and specificity have been achieved by some of the laboratory studies that included subject dependent modeling [20], [24], [27]. Dropped performance has been observed by several studies when implementing subject independent modeling with subject-wise cross validation [24], [25], due to large individual variations in HRV. Several studies applied methods for personalization for detection algorithms to reduce the influence of personal variation [21], [22], [25]. A recent controlled real-road study [25] suggests that the assessment accuracy can suffer from other influential factors on real roads. One study has included an automated driving task with simulated level 2 automated driving [24], however it was not compared with manual driving.

The driving task is one of many confounding factors that can affect HRV during driving. Change of the task demand, such as introducing a secondary task [28], handling complex traffic conditions [29], and driving with driver assistance system [30], may have influence on HR and HRV. Several simulator studies show that compared to manual driving, using advanced driving assistance systems (ADAS) tends to reduce HR [30]–[33], which is an indication of reduced workload [34]. A real-road study also demonstrated lower HR as well as increased HRV during semi-automated driving [35]. Therefore, in order to introduce HRV-based sleepiness detection, it is essential to examine the interaction between sleepiness, automated driving, and HRV.

This paper takes a step forward to study the influence of partial automated driving on the relationship between HRV and sleepiness and investigate the performance of HRV based sleepiness detection in partial automated driving on real roads. In the experiment, a commercial wearable chest band was used to monitor HR. Methods to create personalized assessment are also investigated.

II. Materials and Methods

A. Experiment Setup

1) Participants: This study is a sub-part of a larger study where 89 drivers were recruited, 36 women and 53 men, mean age 38 years (standard deviation [SD] = 11, min 20, max 59 years). There were five inclusion criteria: 1) Participants were required to have experience from ADAS such as adaptive cruise control, lane keep assist and similar; 2) Body mass index below 30, to reduce the risk of sleepiness originating from obstructive sleep disorders); 3) No sleep disorders; 4) No problems with motion sickness; 5) No disabilities preventing the participant from driving an ordinary car. The first criterion was relaxed towards the end of the experiment to be able to recruit the final few drivers. Seventeen drivers did not have any experience with ADAS, whereas 54 drivers had experience with adaptive cruise control, 44 had experience with lane keeping assistance, 48 with parking assistance, and 19 with level 2 assistance. Each participant received 4000 SEK ($\approx$ 400 USD) to compensate for loss of income, due to the long hours needed for participation and recovery sleep on the following day. The study was approved by the Swedish Ethical Review Authority (Dnr 2019-04813) and was also depending on the Swedish government approval of experiments with sleepy drivers on real roads (N2007/5326/TR).

2) Design and Procedure: The study was designed as a within-subject 2 $\times$ 2 design (Fig. 1a), where daytime versus night-time driving and manual versus partially automated level 2 driving were the factors. Each experiment day could accommodate a total of four drivers participating in parallel. Participants drove their first session in the afternoon (daytime or alert condition) and a second session after midnight (night-time or sleep deprived condition). The automated and manual driving sessions took place on two different days. Two cars were used (Volvo V60 and Volvo XC90), drivers were driving the same car on first and second visit. The afternoon driving session started at 3:00 pm (driver A and C) or 5:00 pm (driver B and D), whereas the night drive started at 1:00 am (driver A and C) or 3:00 am (driver B and D). Order was balanced for driving mode (i.e., manual or partial automation), but the alert (daytime) condition always preceded the sleep deprived (night-time) condition. Information material was sent out to participants beforehand, including instructions for the three days before the trials to sleep for at least 7 h each night, go to bed latest at 12:00 midnight, and to get up latest at 9:00 am. They were also asked to fill in a sleep diary covering the three nights leading up to the experiment day, as well as a background information questionnaire.

When arriving to the laboratory, participants received further instructions, concerning both the experiment itself and the test vehicle with its dual command and automation. After reading all information, participants signed an informed consent form. Following the information and consent form, experimenters helped setting up chest bands and sports watches for heart rate measurements and attached electrodes for the physiological reference measurements.

All participants were offered dinner, fruits, non-sugary snacks, water, red tea, or decaffeinated coffee during the evening. The route used was a 90-km section of a dual-lane motorway (road E4, Sweden) where the participants drove from Linköping (exit 111) to Gränna (exit 104) and back. This resulted in an overall driving distance of 180 km, with a speed limit of 120 km/h on the whole section. This road section has, according to the Swedish Transport Administration, an annual average daily traffic of about 8000–15,000 vehicles. In the test vehicle, a test leader was always present, ready to intervene by dual command if the driver showed signs of inappropriate or dangerous driving or was too sleepy to continue. Test leaders

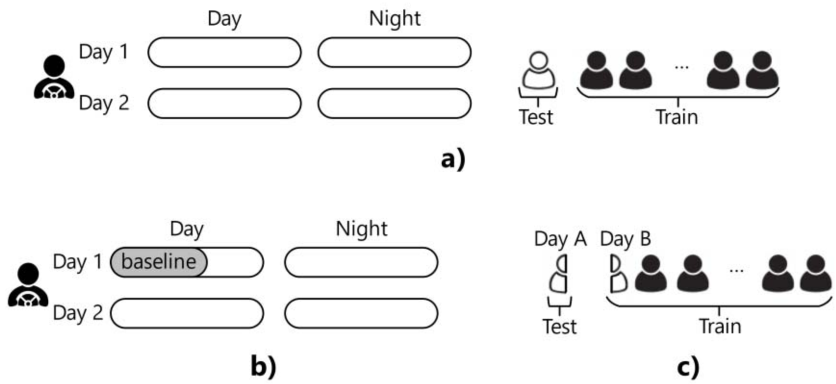

Fig. 1. a) The 2*2 experiment setup for each participant, where they drive in two different days, and in each day they perform two driving's sessions in day- and nighttime, and the leave-one-subject-out cross validation used for the classification model. b) The level 1 personalization uses measurements from start of the first drive during daytime for certain window length to create a HRV baseline. c) the level 2 personalization uses one day measurement in the training set to have a personal calibrated model and test the model with the data from the other day.

did not talk to participants during the data acquisition (except asking for perceived sleepiness, see below).

3) Measurements: Participants were equipped with Garmin brand sports watches (Fénix 5 and Forerunner 645, Garmin Ltd., KS, USA) and Polar H10 chest bands (Polar Electro Oy, Kempele, Finland) for recording RR intervals. H10 has been validated against ECG Holter device, and can provide excellent RR readings under low to moderate intensity activities [36]. Sports watches were used as the data logger for the chest band. On watches, recording interval was set to 'every second' and the option for 'record HRV' was enabled. Other parameters that were measured but not used in this study can be found in [37].

4) Sleepiness Indicators: Every five minutes during data acquisition, participants were prompted by the test leader to verbally report sleepiness according to the Karolinska Sleepiness Scale (KSS) [38]. KSS has nine anchored levels: 1 – extremely alert, 2 – very alert, 3 – alert, 4 – rather alert, 5 – neither alert nor sleepy, 6 – some signs of sleepiness, 7 – sleepy, no effort to stay awake, 8 – sleepy, some effort to stay awake, and 9 – very sleepy, great effort to keep awake, fighting sleep. The reported value should correspond to the drivers average feeling during the previous five min. In this study, KSS $<= 7$ was defined as non-critical condition / alert, KSS > 7 was defined as severe sleepiness. This dichotomization was used by Buendia *et al.* [23] and Persson *et al.* [25].

B. Preprocessing and Features Extraction

For each reported KSS level, five minutes long RR epochs prior to the KSS record were taken for analysis. The outliers of RR intervals were cleaned using the method described in [39]. Time domain, frequency domain, and nonlinear Poincaré plot features were extracted using PhysioNet cardiovascular signal toolbox [40]. Lomb periodogram was used for frequency domain analysis, which does not require resampling for unevenly sampled HRV data. Influence of using different spectral analysis methods and outlier removal methods on the HRV features as well as the relationship between different features and KSS level were discussed in [41] and [23]. Used HRV features are listed in Table I.

TABLE I
LIST OF HRV FEATURES

Domain	Label	Feature
	NN mean	Mean values of NN intervals
	NN variance	Variance of NN intervals
	NN iqr	Interquartile range of NN intervals
Time	**SDNN**	Standard deviation of NN intervals
	RMSSD	Root mean square of successive differences
	pNN50	Percentage of intervals that differ > 50 ms from previous interval
	VLF	Very low frequency spectral power
	LF	Low frequency spectral power
Frequency	**HF**	High frequency spectral power
	LF/HF	Ratio between LF and HF
	total Power	Total spectral power
	SD1	Ellipse width of Poincaré plot
Non-linear	**SD2**	Ellipse length of Poincaré plot
	SD1/SD2	Ratio between SD1 and SD2

Out of the 356 planned trials in the full study, one was cancelled due to technical issues with the logging equipment, two were cancelled due to bad weather, four were cancelled due to hazardous drivers, and 18 were cancelled due to availability and scheduling issues, resulting in 331 trials available for analysis. 54 driving sessions were performed without wearable HR measurement, and were thus excluded from this substudy. In addition, 15 sessions with synchronization problems and four sessions with low-quality wearable HR measures were removed from subsequent analysis. In order to have paired tests among different scenarios, only participants with data from all four driving sessions were kept for analysis, which left 43 participants with 172 driving sessions. In total, the retained dataset contains 3230 5-min epochs. For the included participants, 25 drivers had experience with adaptive cruise control, 12 had experience with lane keeping assistance, 24 with parking assistance, and nine with level 2 assistance.

In addition to HRV features, two other easy-to-access features were also extracted for each epoch, namely duration of driving time and day/night. Day/night is a binary feature separated by 7 pm, which indicates if the driving is performed at late night or not.

All data processing and analysis were performed with Matlab 2020a (MathWorks Inc., MA, USA).

C. Statistical Analysis

All 5-min epochs were grouped by driving conditions (manual and partial automated) and sleepiness (KSS $<= 7$ and KSS > 7). Since the HRV features were not normally distributed, pairwise Mann–Whitney U tests were applied between groups.

The influence of automation and severe sleepiness on HRV metrics were analyzed with mixed-model ANOVA test. The HRV parameters with skewed distribution were logarithmic transformed and a constant was added to features with zero values ($log(x + 1)$). The participant number was set as a random variable and the automation (manual or partial automated driving) and severe sleepiness (KSS $<= 7$ or KSS > 7) were set as fixed variables. Subjects without severe sleepiness episodes were removed from this test to avoid empty subgroups.

For these tests, the level of significance was set at $p < 0.05$ ($p < 0.0125$ after Bonferroni correction for multiple testing).

D. Classification

A machine learning method was employed to assess the possibility of detecting driver sleepiness based on HRV, i.e., distinguishing between KSS > 7 (sleepy) and KSS ≤ 7 (noncritical condition / alert). Considering the high inter-driver variance of HRV levels, leave-one-subject-out (LOSO) cross-validation was used for the training and validation process, which estimates how well the model can generalize for new users. For each iteration, all four driving sessions from the same participant will be held out for testing, and this process is repeated for every participant. To compensate the imbalanced distribution, the severe sleepiness samples were oversampled five times for reaching approximately equal distribution for the two classes. The AdaBoost algorithm was used. The settings used were 50 training cycles with learning rate of 0.1 and a maximum number of splits for each tree set to 30. The training cycles and splits numbers were selected by testing, with higher number of splits, the classifier trended to overfit the training data, and the training cycle number was balanced for training time and performance. The learning rate was used as default value in the Matlab implementation. The area under the receiver operating characteristic (ROC) curve (AUC) was used to measure classification performance. The overall ROC for the cross-validation was plotted by assembling all predictions of test sets in each round. The importance of each feature in the trained ensembled tree of classifiers was examined using Matlab *'predictorImportance'* function. The contribution of each feature was calculated by first summing up the changes in the mean squared error on splits and then dividing by the number of nodes.

TABLE II
NUMBER OF 5-MIN EPOCHS UNDER EACH CONDITION

	Driving Condition		**Daytime**		**Sum**
	Manual	**Auto**	**Day**	**Night**	
KSS<=7	1354	1331	1540	1145	2685
KSS>7	265	280	19	526	545
Sum	1619	1611	1559	1671	3230

To examine if the partial automated driving can influence HRV based sleepiness assessment, specific models that were trained and tested with only manual or automated driving data were developed independently. Their performance was compared with a manual-automated-combined method where the classifier was trained with the whole dataset containing both conditions.

E. Personalization

Two levels of personalization methods were applied to the HRV features. The first level was to add a baseline corrected feature set. The baseline values were derived by using the HRV features from a certain time window at the beginning of the first driving session for the participant (Fig. 1b). The rationale was that the driver was likely to be alert and fully fit to drive at the first driving session at daytime. The baseline HRV level was calculated as the mean value of all 5-min epochs during the selected window size. Then a baseline corrected feature set was generated by normalizing each HRV parameter by dividing all values with the mean baseline level. We tested different window sizes from 5 min to 70 min, as well as the entire driving session. Similar method has been applied by [21] with fixed 3 min window. The second level was to have a personalized calibration by, for each participant, including data from one day in the training session together with all data from other participants and predict the other day when testing (Fig. 1c). This level represented the scenario where a personal calibration with reported KSS is available.

III. RESULTS

A. Sleepy Episodes Under Different Driving Conditions

A summary of the distribution of epochs for different driving conditions and states of sleepiness is shown in Table II. In total, there are 545 out of 3230 5-min epochs with KSS > 7, where most of those epochs (526) occurred in the night driving session. Almost equal number of sleepy epochs can be found with manual (265) and automated (280) driving sessions. At individual level, each participant had 12.67 sleepy epochs in average with a standard deviation of 11.03 epochs.

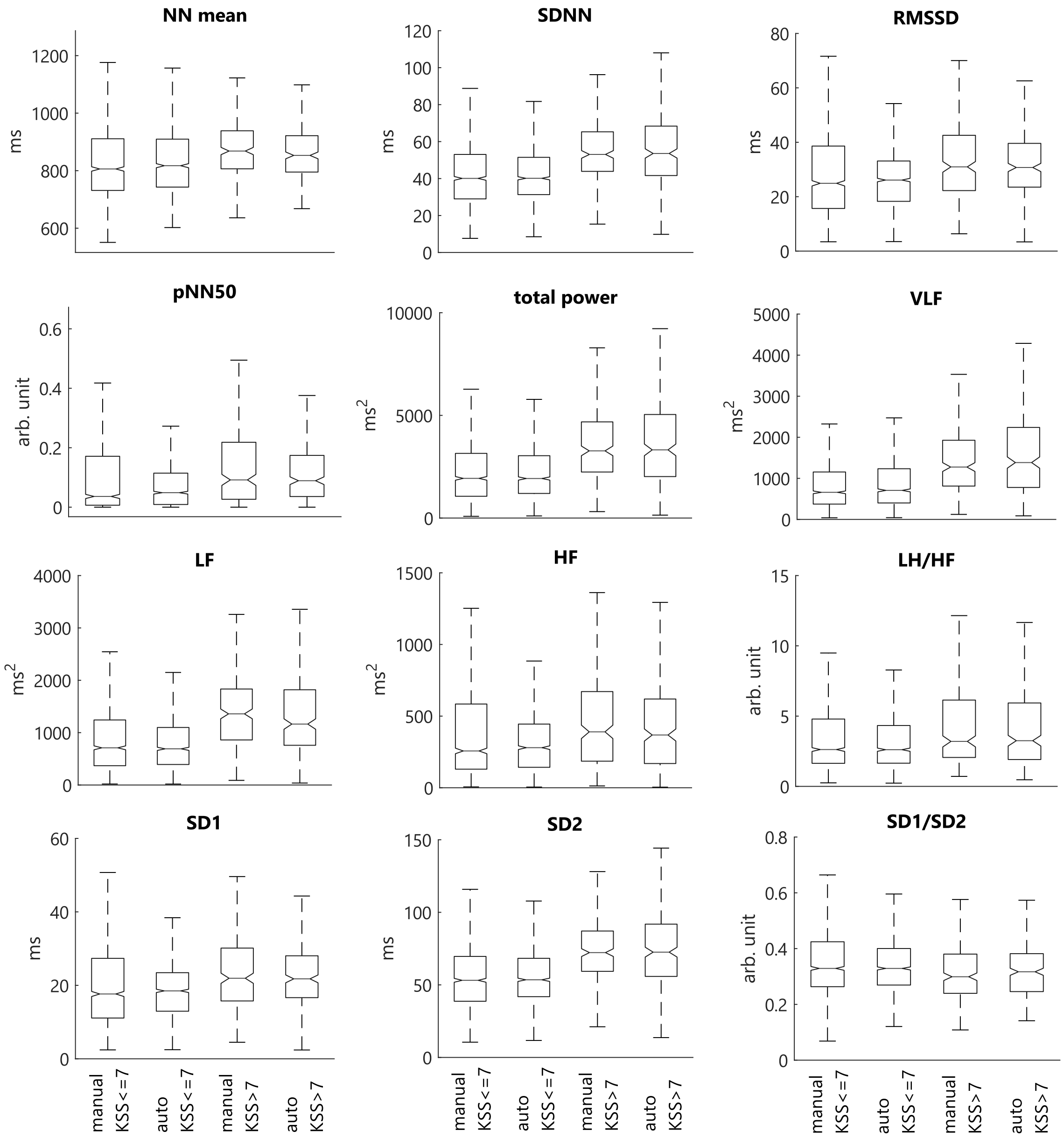

Fig. 2. Box plots for selected HRV features grouped by manual / automated driving and non-critical (KSS <= 7) / sleepy (KSS > 7). Datapoints outside the whiskers (1.5 times the interquartile range above the upper quartile and below the lower quartile) are not shown in the plots.

B. HRV, Automated Driving, and Sleepiness

Boxplots of selected HRV features at group level are shown in Fig. 2. For all HRV parameters, significant differences are found between the KSS > 7 and KSS <= 7 groups for both manual and automated driving conditions. Lower HR, increased SDNN, RMSSD, LF, HF and LF/HF are observed in sleepy episodes. No significant difference is found when comparing manual and automated driving episodes under sleepy or non-critical conditions.

The results of mixed model ANOVA are shown in Table III. When having the participant as a random effect, it shows a significant influence on all HRV parameters. Almost all HRV parameters are significantly affected by severe sleepiness besides pNN50, HF and SD1/SD2, whereas the automation condition shows no significant effect on any feature. For the interaction between severe sleepiness and automation, significant influence can only be found for SD1/SD2.

C. Classification With HRV

Specific models that are trained and evaluated using dataset that contains only manual or partial automated driving episodes have been developed. The performance of the manual specific model is AUC = 0.60, and AUC = 0.68 for the partial automated specific model. No advantage is found for using specific models over the manual-automated-combined (AUC = 0.69), which is developed with the whole data set.

D. HRV Personalization

The classification performance of different calibration procedures is shown in Table IV. Improved performance can

TABLE III
MIXED-MODEL ANOVA RESULTS FOR HRV FEATURES

Features	**Severe Sleepiness (KSS>7)**		**Automation**		**Participant**		**Sleepiness Automation Interaction**	
	F	p	F	p	F	p	F	p
NN mean	**32.46**	**<0.0001**	0.04	0.8516	**6.88**	**<0.0001**	0.66	0.4160
NN variance	**39.69**	**<0.0001**	0.66	0.4213	**7.96**	**<0.0001**	3.68	0.0551
NN iqr	**34.35**	**<0.0001**	1.43	0.2406	**7.13**	**<0.0001**	1.63	0.2017
SDNN	**39.69**	**<0.0001**	0.66	0.4213	**7.96**	**<0.0001**	3.68	0.0551
RMSSD	**9.74**	**0.0035**	0.75	0.3931	**8.23**	**<0.0001**	1.30	0.2551
pNN50	4.76	0.0367	0.56	0.4032	**7.81**	**<0.0001**	2.66	0.2559
VLF	**40.22**	**<0.0001**	2.52	0.1209	**7.78**	**<0.0001**	3.47	0.0626
LF	**23.00**	**<0.0001**	0.01	0.9063	**6.05**	**<0.0001**	0.94	0.3315
HF	3.48	0.0704	0.65	0.4244	**9.25**	**<0.0001**	0.00	0.9498
LF/HF	**10.28**	**0.0028**	1.63	0.2098	**8.39**	**<0.0001**	1.13	0.2875
total Power	**34.69**	**<0.0001**	0.74	0.3952	**8.26**	**<0.0001**	3.99	0.0460
SD1	**9.75**	**0.0035**	0.75	0.3932	**8.23**	**<0.0001**	1.29	0.2553
SD2	**43.16**	**<0.0001**	0.67	0.4192	**7.73**	**<0.0001**	5.39	0.0203
SD1/SD2	5.93	0.0196	0.33	0.5678	**6.75**	**<0.0001**	11.65	0.0007

TABLE IV
CLASSIFICATION PERFORMANCE BY DIFFERENT PERSONALIZATION METHODS

		AUC
No personalization		0.69
+Personal HRV Baseline (Level 1)	5min	0.70
	10 min	0.73
	20 min	0.74
	30 min	0.75
	40 min	0.74
	50 min	0.74
	60 min	0.76
	70 min	0.76
	All	0.75
+Personal calibration (Level 2)		0.80

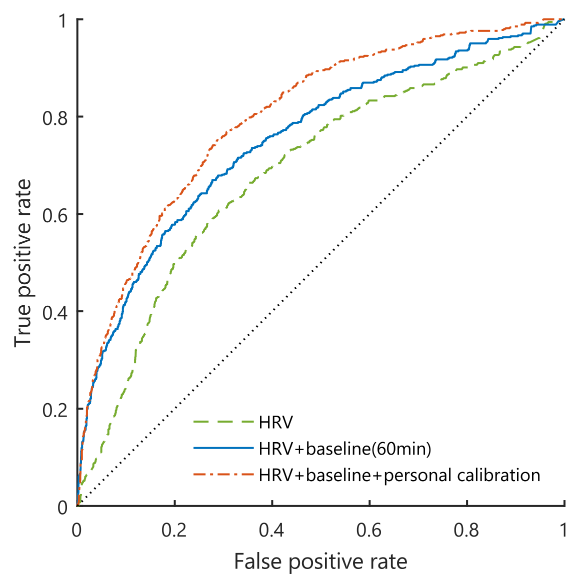

Fig. 3. ROC comparison between non personalization and two levels of personalization methods.

be found for level 1 personalization. When adding baseline corrected feature set, we find longer windows over 20 min bring better performance than shorter windows, for current dataset best performance is achieved with 60 min window baseline. Further improvement is achieved by adding personal calibration for level 2 personalization. The ROC curves of having no personalization, personal baseline and personal calibration are compared in Fig. 3.

E. HRV With Other Features

Fig. 4. shows the performance of using HRV together with other easy to access data. The performance reaches AUC = 0.88 when combing baseline corrected HRV, day/night (AUC = 0.62 when using alone) and driving time (AUC = 0.71 when using alone) together.

F. HRV Feature Importance for Trained Models

The contribution of different HRV features in trained models is shown in Fig. 5 and Fig. 6. For the trained model without personalization, the most important features are NN mean, VLF and SD2 (Fig. 5). When having the baseline corrected HRV feature set, the baseline corrected NN mean becomes the most important feature, and non-personalized NN mean and VLF remains the second and the third (Fig. 6).

IV. DISCUSSION

In this study, the influence of automated driving on the relation between HRV and sleepiness was investigated. Sleepy drivers showed decreased HR, increased HF, LF power and

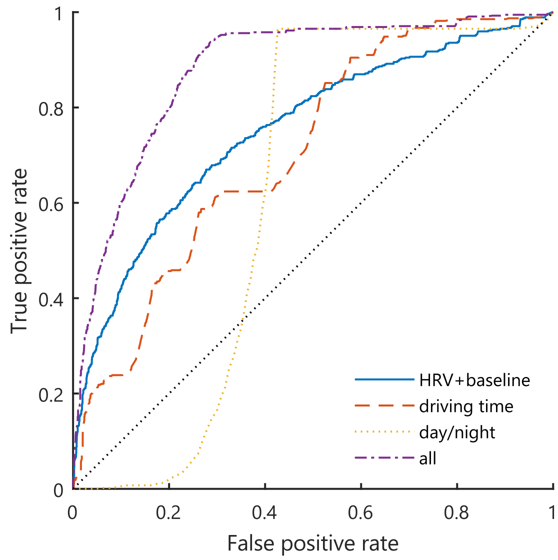

Fig. 4. ROC for models using personalized HRV with baseline correction, driving time, day/night and all three together.

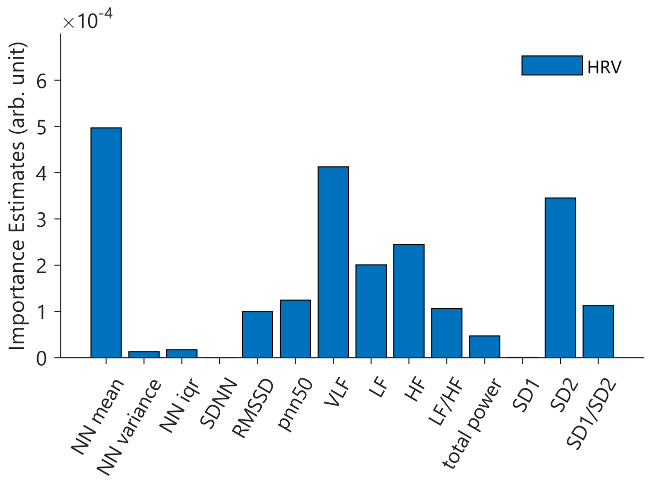

Fig. 5. The feature importance of trained classifier for non-personalized model.

LF/HF for both manual and partial automated driving. The trend is consistent with previous real-road studies with similar setup for manual driving [23], [25]. The trend for HF, LF and LF/HF change is not consistent through all studies [19], [42]. It has been hypothesized that sleepy states activate the parasympathetic nervous system, which leads to higher levels of HF, whereas when the sleep demand is counteracted by subjects fighting to stay awake this will lead to sympathetic activation that increases LF [21]. The inconsistence could also be caused by different experimental setups. Michail *et al.* [42] reported instant LF/HF drop with the occurance of driving error under sleepy state, which is different from studies comparing alert and sleepy state in a longer time span [23], [25]. The fatigue manipulation method and fatigue reference measurement method were not described by Patel *et al.* [19].

Previous studies have shown decreased heart rate when using ADAS. In our study, we did not find significant difference in HR between manual and partial automated driving for the noncritical / alert or severe sleepy group. One possible explanation is the decreased HR using ADAS is in line with the faster sleepiness development. We did not find a significant effect on use of automation for most of the HRV parameters. Similar results have been shown in another real-road study [43], the authors suggesting the reason could be that inherent variability of real-road driving over-shadowed any possible effects resulting from automation. Thus, partial automated driving has minor influence on the relationship between HRV and sleepiness. Furthermore, no benefit was found to train specific HRV based detection model for partial automated driving in the present study.

Due to the high individual variability in HRV, personalization methods have been suggested to improve model performance. Fujiwara *et al.* used a personalized abnormalities detection method [22], whereas Vicente *et al.* used normalized features with baseline established by the first three minutes of driving [21]. Persson *et al.* used baseline established by sections with KSS < 5 [25], this strategy requires labeling measurements with KSS first in practice. In this study, similar method to Vicente *et al.* [21] was applied, where baseline was established by initial driving at daytime. The results show that using averaged baseline over a slightly longer time window (20–60 min) may provide better results. This averaging process may reduce the effect of local HRV fluctuations caused by other factors. However, a too long window will decrease the performance, possibly because the driver state is changing over time. The baseline methods only considered the variation of personal HRV base level but not the range of HRV change. The other method in this study, personal calibration, by collecting one day of labeled data to predict the other day, reached the best performance. However, this method is less practical as it requires labeled data to establish a personal HRV to sleepiness relation.

Different experiment setups, sleepiness labeling methods, validation methods and performance measurements make it difficult to compare results across studies. In the study by Persson *et al.* [25], similar experiment setup was used for manual driving on real roads, but with a three-class classification, for the KSS > 7 group 13.2% sensitivity and 86.8% specificity were reached when using baseline corrected HRV features with LOSO. In comparison, in the present study 49.3% sensitivity was reached at the same (86.8%) specificity with baseline corrected HRV features (Fig. 3). A simulator study [24] compared a wrist-worn HR sensor and ECG for sleepiness detection with 5-min interval, the LOSO result was 78.9% accuracy for ECG and 73.4% for wristband. The performance of HRV with personal baseline in this study was 86.5% accuracy using chest band.

Studies have tried to build multimodal systems to achieve better accuracy, where HRV is used in combination with additional input such as driving performance [44], [45] and other physiological measurements, e.g., EEG and EOG [45]–[47]. The usability can be limited when the input parameters cannot be monitored unobtrusively or are unavailable under automated driving conditions. In this work, good performance was reached by adding day/night and driving-time information,

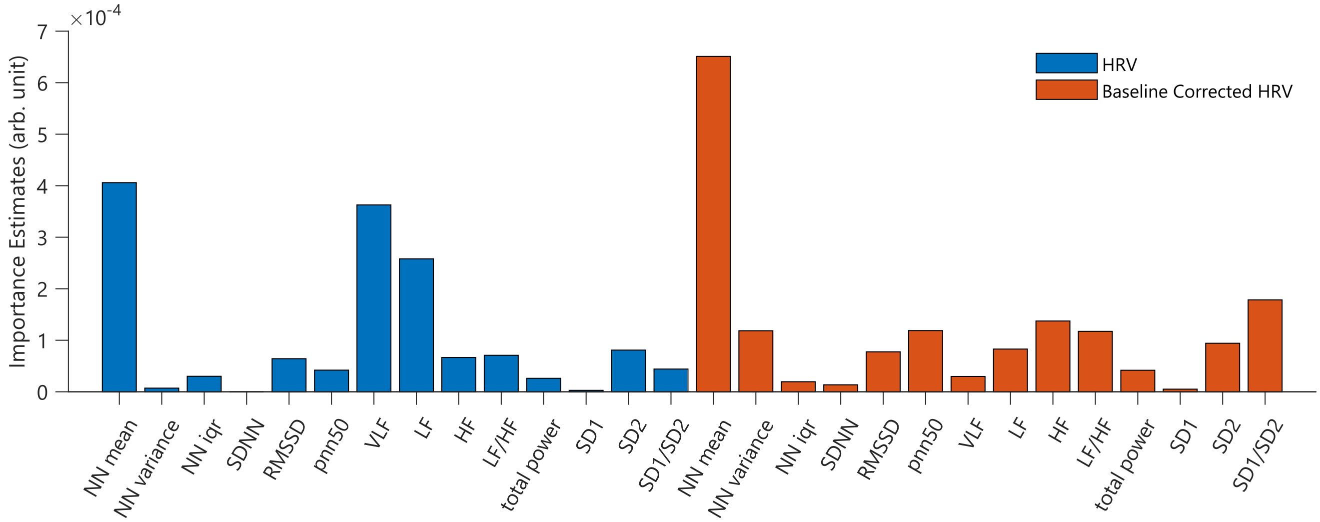

Fig. 6. The feature importance of trained classifier for personalized model with baseline corrected feature set.

which would be simple to retrieve from the wearable device or the vehicle. However, the relation between sleepiness and driving time is related to the driving task. With the same route used in each single trial, this feature may be overfitted to the current experiment. A model using the day/night information may not be suitable for users with different daily schedule such as people with shift works. In future studies, respiration measurement may be a good physiological signal to be used together with HRV [48], [49]. Consumer wearable devices can also provide valuable data beyond driving sessions, such as sleep log, physical activity level, and 24-hour HR measurement, for better estimation.

KSS ratings were used as ground truth data for sleepiness in this study. The KSS has been validated against performance tasks and EEG variables in laboratory settings and it is related to driving performance in simulator and real driving [1], [50], [51]. In this study KSS $= 7$ was selected as threshold for classification as studies have demonstrated that KSS level 8 and 9 are clearly associated with driving performance impairment and physiological changes [1], [52], [53]. Using KSS will not identify exact occurrence of microsleep events, which causes deterioration in driving performance [54], and current 5-min HRV processing window may not be suitable to precisely locate microsleep events that usually have a duration of few seconds. But increase of KSS has been found to correlate with duration and occurrence of microsleep events under sleep restriction [55]. However, there are concerns regarding the validity and reliability in more complex real-life driving situations. In this study and other real-road studies, subjects occasionally report low KSS despite showing obvious signs of sleepiness [56].

The terms sleepiness and fatigue have often been used synonymously in the literature. Efforts have been made to make a clear delineation between fatigue and sleepiness, where some research groups argue that sleepiness is a sub-category of fatigue [57], while others suggest they are distinguishable [58]. Regardless, the clarification of those concepts depends on different causal factors. The major causal factors can be divided into sleep related factors where circadian effects and homeostatic sleep pressure are dominant, and task related factors including time-on-task effect and task load. In this study both sleep and task related factors influenced the subjective sleepiness level. When it comes to the severe sleepiness episodes with KSS above 7, the sleep related factors are assumed to be the major component in our experiment.

Sleep and task related factors are difficult to separate during a driving session. Studies that address those factors separately can be found out of driving context. The relation between HRV and performance of psychomotor vigilance test has been studied under total sleep deprivation for 40 hours [12] and partial sleep deprivation for five nights [14], increased HRV, VLF and LF power have been correlated to decreased vigilance performance. Those findings are in line with our observations in this study. When it comes to task related factors, [59] decreased HR, increased RMSSD, pNN50 and HF were associated to time-on-task effects. However, the change of HRV with time-on-task was not consistent among all studies. Melo *et al.* [60] reported decrease in rMSSD and pNN50 with time-on-task effect. In future studies, the interaction between the sleep and task introduced fatigue and their influence on driving performance and HRV should be investigated.

All four driving tests were conducted on the motorway at fixed time slots. Testing with more sophisticated scenarios that cover different traffic conditions and different schedules for driving and sleeping with extended test groups is required to better reflect real-life driving. Moreover, each subject came for the experiment at two occasions a few days apart, so the current study cannot take care of the intra-individual variation over time. The long-term validity of personalized algorithms needs to be investigated in future studies. The development of driver sleepiness has strong sequential characteristics with time. The current classification model treated each epoch independently, meaning that the information during the transition process to sleep is not fully utilized. New models that consider the time series effect could bring improvements in future studies.

V. Conclusion

This study shows that consumer wearable HR monitors can provide valuable information regarding driver sleepiness under partial automated and manual driving condition. The classification results indicate a potential of using HRV together with other information to form an accurate sleepiness detection system well suited for automated driving scenarios. In our experiment, partial automated driving shows limited impact on the relationship between HRV and sleepiness.

Acknowledgment

The authors acknowledge the contributions by everyone involved in this study, from recruitment of participants, involvement in the experimental planning and preparation of the test vehicles to the data collection: Anna Anund, Mikael Bladlund, Håkan Wilhelmsson, Jonas Ihlström, Ignacio Solis-Marcos, Katja Kircher, Andreas Jansson, Clara Berlin, Marcus Toftin, Simon Duvevik, Gustav Bergholt, David DaSilva Johansson, Gunilla Sörensen, Oscar Järrebring, Martin Zahr, Elsa Magner, Per Gustafsson, Erik Botö and Jawwad Ahmed.

References

[1] D. Hallvig, A. Anund, C. Fors, G. Kecklund, and T. Åkerstedt, "Real driving at night–predicting lane departures from physiological and subjective sleepiness," *Biol. Psychol.*, vol. 101, pp. 18–23, Sep. 2014.

[2] P. Philip and T. Akerstedt, "Transport and industrial safety, how are they affected by sleepiness and sleep restriction?" *Sleep Med. Rev.*, vol. 10, no. 5, pp. 347–356, Oct. 2006.

[3] D. Zwahlen, C. Jackowski, and M. Pfäffli, "Sleepiness, driving, and motor vehicle accidents: A questionnaire-based survey," *J. Forensic Legal Med.*, vol. 44, pp. 183–187, Nov. 2016.

[4] *Euro NCAP 2025 Roadmap: In Pursuit of Vision Zero*, Euro NCAP, Leuven, Belgium, 2017, pp. 1–19.

[5] A. Chowdhury, R. Shankaran, M. Kavakli, and M. M. Haque, "Sensor applications and physiological features in drivers' drowsiness detection: A review," *IEEE Sensors J.*, vol. 18, no. 8, pp. 3055–3067, Apr. 2018.

[6] *Automated Driving—Levels of Driving Automation are Defined in New SAE International*, SAE Standard J3016, SAE, 2016.

[7] M. Körber, A. Cingel, M. Zimmermann, and K. Bengler, "Vigilance decrement and passive fatigue caused by monotony in automated driving," *Proc. Manuf.*, vol. 3, pp. 2403–2409, Oct. 2015.

[8] T. Vogelpohl, M. Kühn, T. Hummel, and M. Vollrath, "Asleep at the automated wheel—Sleepiness and fatigue during highly automated driving," *Accident Anal. Prevention*, vol. 126, pp. 70–84, May 2019.

[9] N. Schömig, V. Hargutt, A. Neukum, I. Petermann-Stock, and I. Othersen, "The interaction between highly automated driving and the development of drowsiness," *Proc. Manuf.*, vol. 3, pp. 6652–6659, Jan. 2015.

[10] J. Gonçalves and K. Bengler, "Driver state monitoring systems–transferable knowledge manual driving to HAD," *Proc. Manuf.*, vol. 3, pp. 3011–3016, Jan. 2015.

[11] J. Wörle, B. Metz, C. Thiele, and G. Weller, "Detecting sleep in drivers during highly automated driving: The potential of physiological parameters," *IET Intell. Transp. Syst.*, vol. 13, no. 8, pp. 1241–1248, Aug. 2019.

[12] E. Chua *et al.*, "Heart rate variability can be used to estimate sleepiness-related decrements in psychomotor vigilance during total sleep deprivation," *Sleep*, vol. 35, no. 3, pp. 325–334, 2012.

[13] K. Kaida, T. Åkerstedt, G. Kecklund, J. P. Nilsson, and J. Axelsson, "Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task," *Ind. Health*, vol. 45, no. 4, pp. 520–526, 2007.

[14] A. Henelius, M. Sallinen, M. Huotilainen, K. Müller, J. Virkkala, and K. Puolamäki, "Heart rate variability for evaluating vigilant attention in partial chronic sleep restriction," *Sleep*, vol. 37, no. 7, pp. 1257–1267, Jul. 2014.

[15] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019.

[16] Y.-L. Zheng *et al.*, "Unobtrusive sensing and wearable devices for health informatics," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 5, pp. 1538–1554, May 2014.

[17] J. R. Pinto, J. S. Cardoso, A. Lourenço, and C. Carreiras, "Towards a continuous biometric system based on ECG signals acquired on the steering wheel," *Sensors*, vol. 17, no. 10, p. 2228, Sep. 2017.

[18] S. Leonhardt, L. Leicht, and D. Teichmann, "Unobtrusive vital sign monitoring in automotive environments—A review," *Sensors*, vol. 18, no. 9, p. 3080, Sep. 2018.

[19] M. Patel, S. K. L. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 7235–7242, 2011.

[20] G. Li and W.-Y. Chung, "Detection of driver drowsiness using wavelet analysis of heart rate variability and a support vector machine classifier," *Sensors*, vol. 13, no. 12, pp. 16494–16511, Dec. 2013.

[21] J. Vicente, P. Laguna, A. Bartra, and R. Bailon, "Drowsiness detection using heart rate variability," *Med. Biol. Eng. Comput.*, vol. 54, no. 6, pp. 927–937, 2016.

[22] K. Fujiwara *et al.*, "Heart rate variability-based driver drowsiness detection and its validation with EEG," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 6, pp. 1769–1778, Nov. 2018.

[23] R. Buendia, F. Forcolin, J. Karlsson, B. A. Sjöqvist, A. Anund, and S. Candefjord, "Deriving heart rate variability indices from cardiac monitoring—An indicator of driver sleepiness," *Traffic Injury Prevention*, vol. 20, no. 3, pp. 249–254, Mar. 2019.

[24] T. Kundinger, N. Sofra, and A. Riener, "Assessment of the potential of wrist-worn wearable sensors for driver drowsiness detection," *Sensors*, vol. 20, no. 4, p. 1029, Feb. 2020.

[25] A. Persson, H. Jonasson, I. Fredriksson, U. Wiklund, and C. Ahlstrom, "Heart rate variability for classification of alert versus sleep deprived drivers in real road driving conditions," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 6, pp. 3316–3325, Jun. 2021.

[26] F. Abtahi, A. Anund, C. Fors, F. Seoane, and K. Lindecrantz, "Association of drivers' sleepiness with heart rate variability: A pilot study with drivers on real roads," in *Proc. IFMBE*, vol. 65, 2017, pp. 149–152.

[27] S. Murugan, J. Selvaraj, and A. Sahayadhas, "Detection and analysis: Driver state with electrocardiogram (ECG)," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 525–537, Jun. 2020.

[28] B. Mehler, B. Reimer, and Y. Wang, "A comparison of heart rate and heart rate variability indices in distinguishing single-task driving and driving under secondary cognitive workload," in *Proc. 6th Int. Driving Symp. Hum. Factors Driver Assessment, Training Vehicle Des.*, 2011, pp. 590–597.

[29] F. Cnossen, T. Rothengatter, and T. Meijman, "Strategic changes in task performance in simulated car driving as an adaptive response to task demands," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 3, no. 3, pp. 123–140, Sep. 2000.

[30] K. A. Brookhuis, C. J. G. van Driel, T. Hof, B. van Arem, and M. Hoedemaeker, "Driving with a congestion assistant; Mental workload and acceptance," *Appl. Ergonom.*, vol. 40, no. 6, pp. 1019–1025, Nov. 2009.

[31] O. Carsten, F. C. Lai, Y. Barnard, A. H. Jamson, and N. Merat, "Control task substitution in semiautomated driving: Does it matter what aspects are automated?" *Hum. Factors*, vol. 54, no. 5, pp. 747–761, Jan. 2012.

[32] D. de Waard, M. van der Hulst, M. Hoedemaeker, and K. A. Brookhuis, "Driver behavior in an emergency situation in the automated highway system," *Transp. Hum. Factors*, vol. 1, no. 1, pp. 67–82, Jan. 1999.

[33] C. J. G. van Driel, M. Hoedemaeker, and B. van Arem, "Impacts of a congestion assistant on driving behaviour and acceptance using a driving simulator," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 10, no. 2, pp. 139–152, Mar. 2007.

[34] J. C. F. de Winter, R. Happee, M. H. Martens, and N. A. Stanton, "Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 27, pp. 196–217, Nov. 2014.

[35] F. N. Biondi, M. Lohani, R. Hopman, S. Mills, J. M. Cooper, and D. L. Strayer, "80 MPH and out-of-the-loop: Effects of real-world semi-automated driving on driver workload and arousal," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, Sep. 2018, vol. 62, no. 1, pp. 1878–1882.

[36] R. Gilgen-Ammann, T. Schweizer, and T. Wyss, "RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise," *Eur. J. Appl. Physiol.*, vol. 119, no. 7, pp. 1525–1532, Jul. 2019.

[37] C. Ahlström, R. Zemblys, H. Jansson, C. Forsberg, J. Karlsson, and A. Anund, "Effects of partially automated driving on the development of driver sleepiness," *Accident Anal. Prevention*, vol. 153, Apr. 2021, Art. no. 106058.

[38] T. Akerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *Int. J. Neurosci.*, vol. 52, nos. 1–2, pp. 29–37, Jul. 1990.

[39] G. D. Clifford, P. E. McSharry, and L. Tarassenko, "Characterizing artefact in the normal human 24-hour RR time series to aid identification and artificial replication of circadian variations in human beat to beat heart rate using a simple threshold," in *Proc. Comput. Cardiol.*, vol. 29, Sep. 2002, pp. 129–132.

[40] A. N. Vest *et al.*, "An open source benchmarked toolbox for cardiovascular waveform and interval analysis," *Physiol. Meas.*, vol. 39, no. 10, Oct. 2018, Art. no. 105004.

[41] F. Forcolin, R. Buendia, S. Candefjord, J. Karlsson, B. A. Sjöqvist, and A. Anund, "Comparison of outlier heartbeat identification and spectral transformation strategies for deriving heart rate variability indices for drivers at different stages of sleepiness," *Traffic Injury Prevention*, vol. 19, pp. S112–S119, Feb. 2018.

[42] E. Michail, A. Kokonozi, I. Chouvarda, and N. Maglaveras, "EEG and HRV markers of sleepiness and loss of control during car driving," in *Proc. 30th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2008, pp. 2566–2569.

[43] J. Stapel, F. A. Mullakkal-Babu, and R. Happee, "Automated driving reduces perceived workload, but monitoring causes higher cognitive load than manual driving," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 60, pp. 590–605, Jan. 2019.

[44] C. Yan, F. Coenen, Y. Yue, X. Yang, and B. Zhang, "Video-based classification of driving behavior using a hierarchical classification system with multiple features," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 30, no. 5, Jun. 2016, Art. no. 1650010.

[45] H. Mårtensson, O. Keelan, and C. Ahlström, "Driver sleepiness classification based on physiological data and driving performance from real road driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 421–430, Feb. 2018.

[46] M. Guo, S. Li, L. Wang, M. Chai, F. Chen, and Y. Wei, "Research on the relationship between reaction ability and mental state for online assessment of driving fatigue," *Int. J. Environ. Res. Public Health*, vol. 13, no. 12, p. 1174, Nov. 2016.

[47] B.-G. Lee, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based driver alertness monitoring with wearable motion and physiological sensor," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 6126–6129.

[48] F. Guede-Fernández, M. Fernández-Chimeno, J. Ramos-Castro, and M. A. García-González, "Driver drowsiness detection based on respiratory signal analysis," *IEEE Access*, vol. 7, pp. 81826–81838, 2019.

[49] S. E. H. Kiashari, A. Nahvi, H. Bakhoda, A. Homayounfard, and M. Tashakori, "Evaluation of driver drowsiness using respiration analysis by thermal imaging on a driving simulator," *Multimedia Tools Appl.*, vol. 79, pp. 17793–17815, Jul. 2020.

[50] K. Kaida *et al.*, "Validation of the Karolinska sleepiness scale against performance and EEG variables," *Clin. Neurophysiol.*, vol. 117, no. 7, pp. 1574–1581, 2006.

[51] T. Åkerstedt, B. Peters, A. Anund, and G. Kecklund, "Impaired alertness and performance driving home from the night shift: A driving simulator study," *J. Sleep Res.*, vol. 14, no. 1, pp. 17–20, 2005.

[52] M. Ingre, T. Åkerstedt, B. Peters, A. Anund, and G. Kecklund, "Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences," *J. Sleep Res.*, vol. 15, no. 1, pp. 47–53, 2006.

[53] C. N. Watling, T. Åkerstedt, G. Kecklund, and A. Anund, "Do repeated rumble strip hits improve driver alertness?" *J. Sleep Res.*, vol. 25, no. 2, pp. 241–247, Apr. 2016.

[54] L. N. Boyle, J. Tippin, A. Paul, and M. Rizzo, "Driver performance in the moments surrounding a microsleep," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 11, no. 2, pp. 126–136, 2008.

[55] C. Bougard *et al.*, "Daytime microsleeps during 7 days of sleep restriction followed by 13 days of sleep recovery in healthy young adults," *Consciousness Cognition*, vol. 61, pp. 1–12, May 2018.

[56] A. Filtness *et al.*, "Bus driver fatigue," Final Rep., 2019.

[57] J. F. May and C. L. Baldwin, "Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies," *Transp. Res. F, Psychol. Behav.*, vol. 12, no. 3, pp. 218–224, 2009.

[58] M. W. Johns, "Drowsy driving and the law," *Tasmania Law Reform Inst.*, vol. 1, p. 12, Nov. 2007.

[59] A. Matuz, D. van der Linden, Z. Kisander, I. Hernádi, K. Kázmér, and A. Csathó, "Enhanced cardiac vagal tone in mental fatigue: Analysis of heart rate variability in time-on-task, recovery, and reactivity," *PLoS ONE*, vol. 16, no. 3, 2021, Art. no. e0238670.

[60] H. M. Melo, L. M. Nascimento, and E. Takase, "Mental fatigue and heart rate variability (HRV): The time-on-task effect," *Psychol. Neurosci.*, vol. 10, no. 4, p. 428, Dec. 2017.

Ke Lu (Member, IEEE) received the M.Sc. degree in Electronics from the University of Edinburgh, U.K., in 2014, and the Ph.D. degree in Biomedical Engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2019. He is currently a Post-Doctoral Researcher in Biomedical Signals and Systems with the Department of Electrical Engineering, Chalmers University of Technology. His research interests include physiological measurement and biomedical signal processing.

Johan Karlsson received the M.Sc. degree in Cognitive Science from the Computer Science Department, Linköping University, Linköping, Sweden, in 2002. His research interests include driver support and driver state assessment within the light vehicle driving environment, with particular focus on driver sleepiness.

Anna Sjörs Dahlman received the M.Sc. degree in Engineering Biology in 2006 and the Ph.D. degree in Rehabilitation Medicine in 2010. She is currently an Adjunct Associate Professor with the Department of Biomedical Signals and Systems, Chalmers University of Technology, and also a Researcher with the Swedish National Road and Transport Research Institute. Her research interests include physiological monitoring and driver health.

Bengt Arne Sjöqvist received the M.Sc. degree in Electrical Engineering in 1976 and the Ph.D. degree in Applied Electronics/Biomedical Engineering in 1984. He was an Associate Professor in 1989. He is currently a Professor of practice emeritus digital health in the Biomedical Signals and Systems group with the Chalmers University of Technology. His research interests are within biomedical engineering, digital health, and traffic safety, with focus on prehospital and distance care, including decision support systems.

Stefan Candefjord received the M.Sc. degree in Engineering Physics in 2006 and the Ph.D. degree in Biomedical Engineering in 2011. He is currently an Associate Professor with the Biomedical Signals and Systems, Department of Electrical Engineering, Chalmers University of Technology. His research interests are within biomedical engineering, digital health and traffic safety, with focus on decision support systems.