Driver Sleepiness Classification Based on Physiological Data and Driving Performance From Real Road Driving

Henrik Mårtensson, Oliver Keelan, and Christer Ahlström

***Abstract*— The objective of this paper is to investigate if signal analysis and machine learning can be used to develop an accurate sleepiness warning system. The developed system was trained using the supposedly most reliable sleepiness indicators available, extracted from electroencephalography, electrocardiography, electrooculography, and driving performance data (steering behavior and lane positioning). Sequential forward floating selection was used to select the most descriptive features, and five different classifiers were tested. A unique data set with 86 drivers, obtained while driving on real roads in real traffic, both in alert and sleep deprived conditions, was used to train and test the classifiers. Subjective ratings using the Karolinska sleepiness scale (KSS) was used to split the data as sufficiently alert (KSS $\leq$ 6) or sleepy (KSS $\geq$ 8). KSS = 7 was excluded to get a clearer distinction between the groups. A random forest classifier was found to be the most robust classifier with an accuracy of 94.1% (sensitivity 86.5%, specificity 95.7%). The results further showed the importance of personalizing a sleepiness detection system. When testing the classifier on data from a person that it had not been trained on, the sensitivity dropped to 41.4%. One way to improve the sensitivity was to add a biomathematical model of sleepiness amongst the features, which increased the sensitivity to 66.2% for participant-independent classification. Future works include taking contextual features into account, using classifiers that takes full advantage of sequential data, and to develop models that adapt to individual drivers.**

***Index Terms*— Driver sleepiness, feature selection, classification, real driving.**

I. Introduction

MANY drivers experience driver sleepiness [1]–[4], a condition that causes crashes and results in injuries and fatalities [5], [6]. According to the world health organization, 1.25 million people are killed in road crashes every year, and amongst people aged 15–29, it is the number one cause of death [7]. The proportion of these crashes that are due to sleepiness is about 10–20 % [8]–[10].

Drivers are usually aware of being sleepy [11], [12], but many continue to drive nonetheless. To nudge drivers to stop driving, several intelligent sleepiness warning systems have been introduced and some systems are already on the market. As a prerequisite for compliance and acceptance, these warning systems must be able to accurately monitor the driver's condition. Unfortunately, the quantification of sleepiness remains a challenge, and a solid objective measure of sleepiness has yet to be found [13].

Sleepiness detection systems are based on (i) vehicle-based information, (ii) behavioural information and/or (iii) physiological information [14]–[16]. Vehicle-based measures, such as steering wheel activity and lane positioning, are already available in modern vehicles and have the advantage of being nonintrusive. However, the warnings are often inaccurate and unreliable [16]. Behavioural measures, such as gaze behaviour and eye closures, can be camera-based and are thus nonintrusive, but the extraction of eye features from video data is still difficult, especially in quickly changing light conditions [17]. Physiological measures have been found to be more reliable [16], but they are as of yet very obtrusive. A fourth information source, that in our opinion has been underexploited so far, involves biomathematical modelling of sleepiness. Such models make use of information about the time of day, the time since awakening and the duration of prior sleep [18], [19]. Biomathematical models estimate latent sleepiness, which is the fundamental sleepiness that is regulated by homeostatic mechanisms and the circadian system. As such, biomathematical models are a great complement to the more direct indicators mentioned above.

When designing sleepiness classification systems based on physiological data, the main source signals are electroencephalography (EEG) to measure brain activity, electrooculography (EOG) to measure blink behaviour, and electrocardiography (ECG) to measure heart rate and heart rate variability (HRV). The EEG data are commonly quantified using the power in the theta (4–7 Hz), alpha (8–15 Hz) and beta (16–31 Hz) frequency bands. An increased theta power reflects sleep need [20], [21], but in a driving setting, increased alpha power appears to be a more reliable indicator [22], [23]. Heart rate, HRV and respiration have also been used to quantify driver sleepiness [24]–[29].

Manuscript received September 19, 2017; revised January 8, 2018 and February 28, 2018; accepted March 2, 2018. Date of publication April 4, 2018; date of current version January 31, 2019. This work was supported in part by the Swedish Innovation Agency through the FFI Program and in part by the Competence Centre Virtual Prototyping and Assessment by Simulation. The Associate Editor for this paper was S. E. Li. *(Corresponding author: Christer Ahlström.)*

H. Mårtensson was with the Department of Biomedical Engineering, Linköping University, 58185 Linköping, Sweden. He is now with Saab Aeronautics AB, 58188 Linköping, Sweden.

O. Keelan was with the Department of Biomedical Engineering, Linköping University, 58185 Linköping, Sweden. He is now with ESSIQ AB, 42150 Västra Frölunda, Sweden.

C. Ahlström is with the Swedish National Road and Transport Research Institute, S-58195 Linköping, Sweden (e-mail: christer.ahlstrom@vti.se).

Digital Object Identifier 10.1109/TITS.2018.2814207

The parasympathetic influence when falling asleep slows down the heart and make its beating less regular [30], [31]. However, there are many confounding factors and the HRV results are therefore often ambiguous. Finally, EOG data are typically quantified in terms of blink durations or eyelid opening/closing velocities eg. [32], [33]. Numerous signal analysis methods have been used to extract features that capture these properties in the raw signals, including phase synchronization [34], spectral analysis [35]–[39], joint time/frequency and Wavelet transforms [40]–[42], and nonlinear approaches such as fractal dimensions and different entropies [36], [43], [44]. These features are then often used as input to various machine learning algorithms, such as neural networks [27], [45], support vector machines (SVM) [36]–[38], [40], [41], [46] and hidden Markov models [28], [35], [42].

There are three major shortcomings that many driver sleepiness classification studies suffer from. *First*, the results are often based on rather small datasets with 5–20 participants. This small amount of data does not allow for proper validation of the sophistic methods that are used e.g. [27], [34], [35], [40], [43]–[48]. *Second*, the data has often been acquired in driving simulators and not from real road driving e.g. [27], [28], [34], [36], [37]–[41], [43], [45]–[49]. One important limitation of using driving simulator data is that the drivers do not perceive any risk, which may cause a behaviour that is different from that on real roads [50]. *Third*, many studies are not actually using data from sleepy drivers. Instead "sleepy" data is invoked by driving for about an hour in a monotonous setting e.g. [34], [35], [38], [41], [43], [45], [46], [48]. Such an experimental design gives rise to fatigue caused by under-stimulation rather than to physiological sleepiness [51]. This paper is based on data from three different field studies with a total of 86 participants driving on real roads in real traffic. As far as we know, this constitutes the largest real road driver sleepiness dataset in the world, containing, by design, data from both alert and sleep deprived conditions.

The objective of this paper is to investigate if signal analysis and machine learning can be used to develop an accurate sleepiness warning system. The proposed system will be developed within a standard machine learning framework, including feature extraction, feature selection, classifier selection and classifier validation. Secondary aims are to investigate which features are most informative for sleepiness detection, the potential added value of a biomathematical model when classifying sleepiness, and the impact of individual differences on detection performance.

II. Sleepiness Database

The database used in this paper consist of data from three separate driver sleepiness experiments. The first experiment involved 18 drivers (8 women, mean age 41 years) who drove for about 90 minutes on a motorway. Each driver drove two times, once in a supposedly alert state during daytime and once in a sleep deprived state during night-time. The second experiment included 24 drivers (12 women, mean age 35 years) who drove for about 135 minutes three times on a motorway (supposedly alert during daytime, mostly alert during the evening and sleep deprived during night-time). The third experiment included 44 drivers (21 women, mean age 44 years) who drove for about 90 minutes three times on a rural road (supposedly alert during daytime, mostly alert during the evening and sleep deprived during night-time). The participants were recruited by random selection from the Swedish register of vehicle owners. Each of the three experiments were approved by the Regional Ethics Committee in Linköping, Sweden, and in addition, the Swedish government approved running tests with sleepy drivers on real roads (N2007/5326/TR). The experimental car was equipped with dual command at the front right passenger seat, allowing the test leader to take control of the vehicle if necessary.

Physiological data (EEG, EOG and ECG) were acquired by a portable digital recording system (Vitaport 2 and 3, Temec Instruments BV, the Netherlands) that was synchronized with the vehicle's logging equipment. The electrodes used for EOG (measured vertically and horizontally across the eyes) and ECG (lead II) were of the disposable Ag/AgCl type. The EEG was measured via three bipolar derivations positioned at Fz-A1, Cz-A2 and Oz-Pz using silver plated non-disposable electrodes. The sampling frequency was 512 Hz for the EOG and 256 Hz for the EEG and ECG. Vehicle-data from the car, such as speed, steering wheel angle, yaw rate, acceleration, etc., were logged with a sampling frequency of 40 Hz. In addition, the vehicle was equipped with a GPS receiver and a lane tracker (Mobileye Vision Technology Ltd., Israel). The Karolinska Sleepiness Scale (KSS) was used to acquire self-reported sleepiness every fifth minute during the drives. KSS has nine anchored levels [52]: 1–extremely alert, 3–alert, 5–neither alert nor sleepy, 7–sleepy, no effort to stay awake, and 9–very sleepy, great effort to keep awake, fighting sleep. The reported value corresponds to the average feeling during the past 5 minutes. The KSS values are used as the target values when training the machine learning algorithms.

All participants were prepared in the same way in all experiments. Exclusion criteria were shift workers, participants who had travelled across at least three time zones during the past two weeks, participants with sleep or health problems, and participants who drove less than 5000 km/year. The participants filled out sleep/wake diaries during the three days before the start of the study. When filling out the sleep/wake diaries the participants also practiced using KSS. Before arrival, the participants were requested to avoid alcohol for 72 hours and to abstain from nicotine and caffeine for 3h before driving. More detailed information about the three experiments can be found in [53]–[55].

III. Methods

The machine learning pipeline used in this work is summarized in Fig 1. The raw signals were filtered and divided into 2.5-minute epochs in a pre-processing stage. For each epoch, 54 different features were calculated. Sequential Forward Floating Feature Selection (SFFS), taking inter-feature dependencies into account, was applied to reduce the dimensionality of the dataset. The selected features were then used to classify the data as either severely sleepy (KSS $\geq$ 8) or as sufficiently alert (KSS $\leq$ 6). To obtain a clear separation of the two

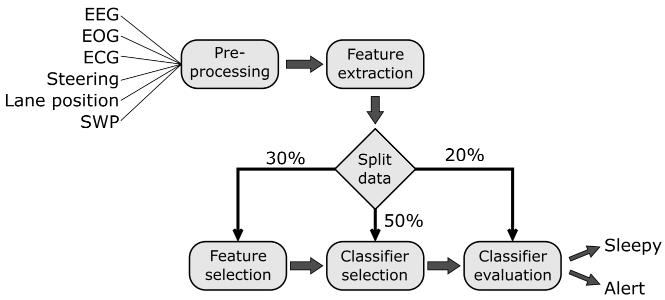

Fig. 1. The machine learning pipeline, going from raw data, via classifier design to decision.

classes, samples with KSS = 7 were discarded as outlined by Sandberg *et al.* [19]. The class definitions are motivated by the fact that KSS levels 8 and 9 are associated with severe signs of physiological sleepiness and increased frequency of lane-departure incidents [11], [52].

The dataset was split up into a feature selection set (30 %), a training set (50 %) and a test set (20 %) using a holdout approach. The idea with the three datasets was to avoid data leakage between the different phases of the machine learning pipeline. Data were distributed randomly between the three datasets. All data processing was carried out in Matlab version 9.2 (The Mathworks Inc., Natick, MA, USA).

A. Pre-Processing

The EOG signals were lowpass filtered with a 5th order Butterworth filter at 11 Hz, and the ECG signal was bandpass filtered with a 5th order Butterworth filter with cut-off frequencies at 0.5 and 45 Hz. The EEG data were bandpass filtered with a Chebyshev Type II filter using a lower cut-off frequency of 4 Hz and an upper cut-off at 30 Hz. EEG data with amplitudes above three times the standard deviation were considered as artifacts and were rejected from further analyses. This is a very simple artifact handling procedure, and the aim was simply to remove the worst outliers from the data. All filtering was conducted with zero-phase forward and reverse digital IIR filtering.

B. Feature Extraction

Each feature was calculated based on 2.5-minute segments of the signals. Since the KSS ratings were given every fifth minute, this provides two feature values per KSS rating. The 2.5-minute segment duration was chosen as a compromise between having enough raw data to calculate the features, and short enough to give a reasonably prompt sleepiness detection.

Eight frequency domain features were extracted from each of the three EEG channels [56]: the absolute power in the θ and α-bands, $\theta/(\theta+\alpha)$, $\alpha/(\theta+\alpha)$, $(\theta+\alpha)/\beta$, α/β, $(\theta+\alpha)/(\theta+\beta)$ and θ/β. In addition, two nonlinear EEG features were extracted per channel. Sample entropy [57], calculated with dimension = 2 and threshold = 0.2 · std(EEG), and the Higuchi fractal dimension [58], calculated with kmax = 70. In total, 30 EEG features were extracted from the database.

Four HRV features were extracted from the ECG [24]–[28]: LF/HF, HF/(LF + HF), LF/(LF + HF) and the root mean square of successive differences (RMSSD). HRV was derived as the time difference from the interbeat interval signal, and LF and HF were calculated as the frequency power in the 0.04–0.15 Hz range (LF) and in the 0.15–0.4 Hz range (HF). In addition, sample entropy (with dimension = 2 and threshold = 0.2 · std(ECG)) and Higuchi dimension (with kmax = 70) were extracted from the raw ECG signal. In total, 6 ECG features were extracted.

Eight blink related features were derived from the vertical EOG: the mean and the 90th percentile of the blink duration, the blink amplitude, the eyelid closure speed and the eyelid opening speed. The blink parameters were extracted from the vertical EOG signal with an automatic blink detection algorithm based on derivatives and thresholding [59]. To reduce problems with concurrence of eye movements and blinks, the blink duration was calculated at half the amplitude of the upswing and the downswing of each blink and defined as the time elapsed between the two. In addition, sample entropy (with dimension = 2 and threshold = 0.2 · std(EOG)) and Higuchi dimension (with kmax = 6) were extracted from the raw EOG signal. In total, 10 EOG features were extracted.

Seven features based on vehicle dynamics were extracted to capture deteriorating driver performance. The standard deviation of steering wheel angle, the rate of micro steering corrections, the rate of large steering corrections, the standard deviation of lane position, the standard deviation of lateral accelerations, and Hjorth's second order descriptor [60] of lane position and lateral acceleration. The intention of these features was to capture when the driver was either too inactive to maintain a straight course, and when the driver was making large steering corrections (because of prior inactivity) [19], [61].

A mathematical model of sleepiness, the so called sleep/wake predictor (SWP) was also included as a feature [62]. SWP is calculated as SWP = 10.9 − 0.6(S + C + U), where S is the level of alertness which depends on time awake and prior sleep, C is a circadian component, and U represents the ultradian rhythm. In this paper, where the participants followed a prescribed sleep agenda, it was assumed that the participant slept between 23.00-07.00 the night before the trials.

C. Feature Selection

SFFS based on decision tree classification was used to reduce the dimensionality of the feature set [63], [64]. The stopping condition was when adding more features did not improve the overall optimisation score, 2^(sensitivity · specificity). To increase the certainty of the feature selection stage, the SFFS algorithm was evaluated over 15 iterations, and in each iteration a 10-fold cross validation was performed. The final feature set was selected using the candidate set with the highest optimisation score. SFFS was based on the feature selection dataset, comprising 30 % of the full dataset.

SFFS was applied to two different feature sets, the first with all 54 features, and the second with all features except SWP. SWP has previously been shown to boost the results when classifying sleepiness based on driving performance features [19]. However, SWP requires that the driver provides

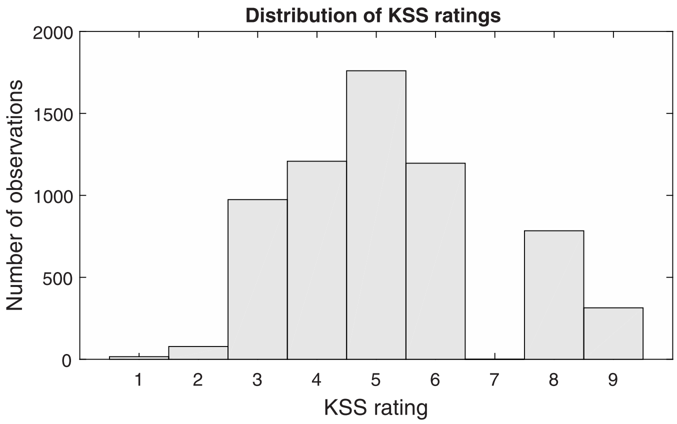

Fig. 2. Distribution of KSS ratings in the dataset. Note that KSS = 7 has been removed, see the main text for details.

information about the time since awakening and the duration of prior sleep. By excluding SWP, we aimed to get a better picture of (i) the performance we can expect from a fully automatic stand-alone system, and (ii), which features that carries roughly the same information as SWP.

D. Classifier Design and Selection

The full feature matrix consisted of 6330 observations and 54 features. In total, there were 5232 cases in the alert class and 1098 cases in the sleepy class. The observations were standardized by Standard score normalisation.

Five different classifiers were evaluated: k-nearest neighbours (kNN), linear SVM, Gaussian SVM, AdaBoost and random forest. These classifiers were chosen because they are well established and because there is a clear difference in complexity and computational cost between them. The *AdaBoost* classifier used decision trees as weak learners with a maximum number of decision splits set to 20 and a learning rate of 0.1. The AdaBoost.M1 algorithm was used as ensemble aggregation method. The *random forest* also used decision trees as weak learners and applied bootstrap aggregation, or bagging, as the ensemble-aggregation method. The *linear SVM* and the *Gaussian SVM* used a heuristic procedure to set an appropriate kernel scale factor. The *kNN* used a Euclidean distance function with no distance weighting. The number of iterations, number of trees, number of neighbours and the error penalty parameter, respectively, were selected by evaluating the achieved sensitivity for a range of parameter values. All classifiers were trained using 10-fold cross validation. Since the distribution of alert versus sleepy cases was skewed (Fig 2), a cost function with a penalty of 5 for erroneously classified sleepy observations was used. Remaining classifier parameters were set to their default values in the Statistics and Machine Learning Toolbox version 11.1 (The Mathworks Inc., Natick, MA, USA). Both parameter optimisation and classifier selection were based on data from the training set. Since correct sleepy classifications were deemed more important than misclassified alertness, the classifier with the best sensitivity results and the most robust behaviour across parameter values was selected as the final classifier.

TABLE I

RESULTING FEATURE SETS AFTER RUNNING SFFS. THE LEFT COLUMN LISTS THE MOST DISCRIMINATING FEATURES WHEN SWP IS EXCLUDED FROM THE FULL FEATURE SET, AND THE RIGHT COLUMN LISTS THE RESULTING FEATURE SUBSET AMONGST ALL FEATURE CANDIDATES

SWP excluded	SWP included
Mean blink duration	SWP
Higuchi EEG (Oz-Pz)	θ/β EEG (Cz-A2)
Sample entropy ECG	Mean Lid Closure Speed
90^{th} %-tile lid closure speed	90^{th} %-tile blink duration
θ power EEG (Fz-A1)	Higuchi ECG
α/(θ+α) EEG (Fz-A1)	α/β EEG (Oz-Pz)
θ/β EEG (Cz-A2)	θ/(θ+α) EEG (Fz-A1)
θ/(θ+α) EEG (Cz-A2)	RMSSD
RMSSD	α power EEG (Cz-A2)
Mean lid opening speed	α power EEG (Fz-A1)
Blink frequency	
α power EEG (Cz-A2)	
90^{th} %-tile blink duration	
Sample entropy EOG	

E. Evaluation

The final classifier was evaluated on the test dataset in terms of accuracy, sensitivity and specificity. Note that the selection of the final classifier was done based on the training dataset, and that only the final classifier was evaluated on the test dataset. This was to prevent data leakage from the classifier selection step.

In addition, a leave one out evaluation was performed to estimate how generalisable the final classifier was to unseen drivers. In this approach, one participant was held out as test data and the rest of the participants were used as training data. This procedure was repeated for all participants, and the mean accuracy, sensitivity and specificity across participants was calculated.

IV. RESULTS

A. Feature Selection

The feature selection process showed that performance increased rapidly after including the first few features, see Fig 3. In the case where SWP was excluded from the candidate features, the increase in performance saturated after about 6–7 features and reached an optimum at 14 features. These final 14 features are listed in Table I. In the case where SWP was included amongst the candidate features, the optimisation score jumped to 93 % as soon as SWP entered the model (in iteration 8, Fig 3). An optimum was reached when 10 features were selected.

B. Classifier Selection

Accuracy, sensitivity and specificity for different parameter values for the five candidate classifiers are shown in Fig 4. For the random forest classifier, the results stabilized after including about 10 voting trees (both for the feature sets excluding and including SWP). For the AdaBoost classifier, the accuracy was stable throughout the iterations. However, the sensitivity increased and the specificity decreased with the number of

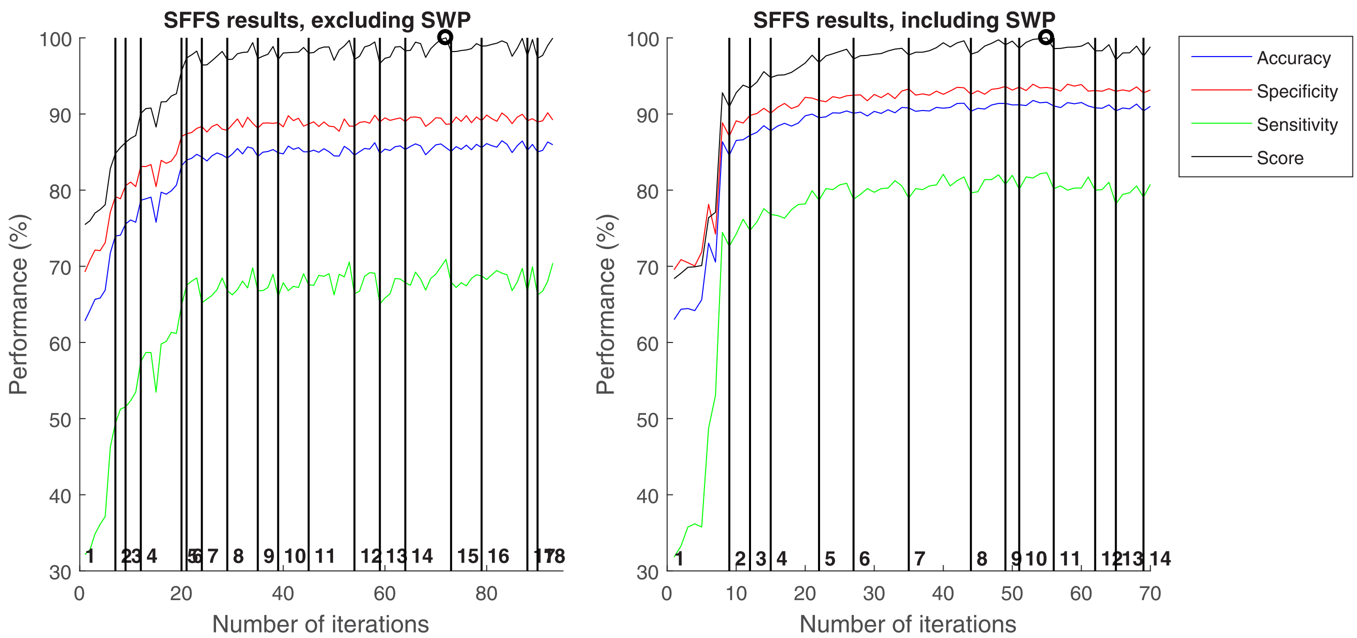

Fig. 3. Evolution of accuracy, sensitivity, specificity and optimization score when running SFFS. Each vertical line marks when the number of features (also written with bold numbers in the lower part of the figure) in the so far best performing feature set changes. The subset with the best optimization score is marked with a ring. SWP is excluded from the set of feature candidates in the left subplot and included in the right subplot.

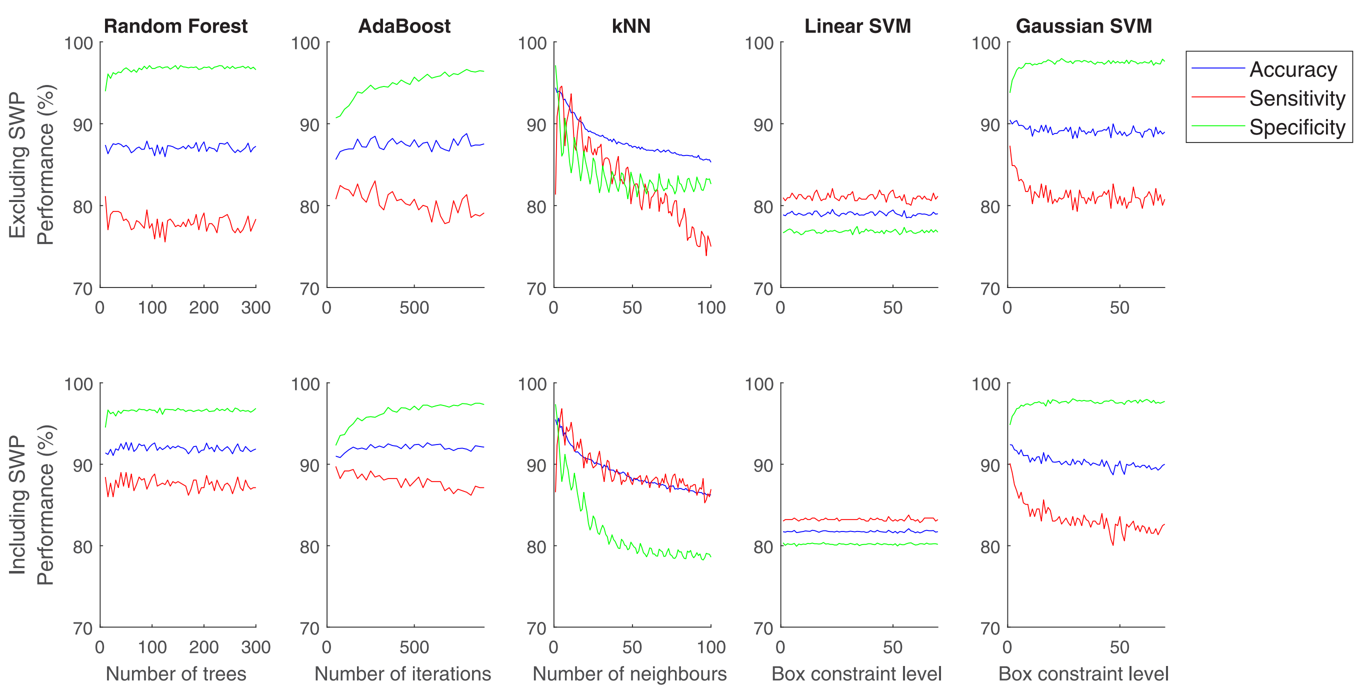

Fig. 4. Accuracy, sensitivity and specificity for the five candidate classifiers.

weak learners. For kNN, there were large fluctuations in both the sensitivity and the specificity with increasing numbers of neighbours. For Gaussian SVM, performance stabilized at a box constraint level of about 10.

The performance was similar for random forest, AdaBoost and Gaussian SVM. However, we chose random forest to be the final classifier since it (i) provided stable results on the training set across parameter values, (ii) performed well on both feature sets, and (iii) is known to be robust to overfitting and to reduce the variance. It was decided to use 55 voting trees since the sensitivity values stabilized around this value.

C. Evaluation of Final Classifer

Receiver operating characteristic (ROC) curves for the selected random forest classifier, based on the feature sets with and without SWP, are shown in Fig 5. Accuracy, sensitivity and specificity are reported in Table II. In general, the accuracy was similar between the two feature sets (93.5 % without SWP, 94.1 % with SWP), but the sensitivity was

TABLE II
ACCURACY, SENSITIVITY AND SPECIFICITY FOR THE FINAL CLASSIFIER BOTH WHEN EXCLUDING AND INCLUDING SWP IN THE FEATURE SET

		Accuracy	Sensitivity	Specificity
Excluding SWP	Training set	91.1	78.2	96.3
	Test set	93.5	80.3	96.3
	Leave one participant out	84.0	41.4	93.1
	All data	96.5	89.8	98.0
Including SWP	Training set	91.7	87.1	96.6
	Test set	94.1	86.5	95.7
	Leave one participant out	86.9	66.2	93.3
	All data	97.2	95.7	97.9

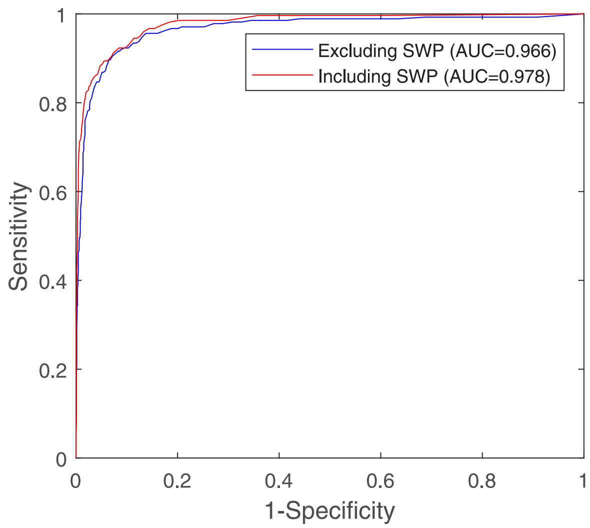

Fig. 5. ROC curves for the two final random forest classifiers, when SWP is included or not.

higher when SWP was included (80.3 % versus 86.5 %). For participant-independent classification (leave one participant out), sensitivity dropped considerably. Finally, when the classifier was trained on all available data, performance increased in comparison with the training set results.

To make sure that the obtained test results were not only valid for this one randomized split of the dataset, the entire procedure was repeated 10 times with different randomized partitions. In this evaluation, the feature selection step was redone and the random forest classifier was retrained in each iteration. The same classifier parameters (random forest with bootstrap aggregation and 55 decision trees) were used in all repetitions though. This resulted in a mean $\pm$ std accuracy of 93.8 $\pm$ 0.6 (range 93.1–94.7), a mean $\pm$ std sensitivity of 80.6 $\pm$ 3.7 (range 76.7–88.1), and a mean $\pm$ std specificity of 96.6 $\pm$ 0.7 (range 95.8–97.6) when SWP was excluded. Corresponding results when SWP was included were a mean $\pm$ std accuracy of 95.0 $\pm$ 0.3 (range 94.4–95.4), a mean $\pm$ std sensitivity of 87.8 $\pm$ 2.3 (range 84.1–91.3), and a mean $\pm$ std specificity of 96.5 $\pm$ 0.6 (range 95.3–97.3).

V. DISCUSSION

A driver sleepiness detection system based on a random forest classifier, trained and tested with data obtained from real roads in real traffic, has been developed. The classification accuracy on the test dataset was 94.1 %. By incorporating a biomathematical model of sleepiness amongst the features, a 6.2 % increase in sensitivity was achieved. The importance of personalized sleepiness detection systems was also demonstrated. When testing the classifier on data from a person that it has not been trained on, the sensitivity dropped with about 20 %.

Several interesting observations can be made regarding the selected feature sets. *First*, it is noteworthy that none of the selected features in Table I are based on driving performance. This indicates that data from the vehicle, such as steering wheel activity and lane position, are less accurate in classifying sleepiness compared to physiological data. This is especially interesting since most fatigue detection systems included in production vehicles are based solely on driving performance data. Adding blink related information from a remote eye tracker or heart rate data from sensors in the steering wheel or seat would greatly improve the accuracy of those systems. A clear practical drawback with physiological data is however that this requires individualised or trait-based training data, see further the section on individual differences below. *Second*, the benefit of including information about prior sleep, time awake and time of day is clearly demonstrated in Fig 3. When SWP entered the feature set, the accuracy immediately increased to 86 %. This is a known phenomenon [19]. The downside of using SWP is that it only works on a population level, but not so well on an individual level [62]. This is where machine learning has an important role to play. By adding more information, SWP can be both personalized and customized to the situation at hand. *Third*, when SWP was excluded from the candidate features, the mean blink duration took its place as the most contributing feature, reaching an accuracy of 74 % (Table I). The effect size of blink durations is known to be small, but increased blink durations has been found for increasing levels of driver sleepiness in essentially all studies where it has been measured eg. [32], [33]. Our results, in combination with previous findings, indicate that SWP and mean blink duration play similar roles in sleepiness classification, i.e. as general measures that provide stable results on a group level, but that may fail on an individual level or in a particular situation. Adding more features improves the results, likely by pushing the general SWP (or blink duration) estimate in the right direction. *Forth*, physiological data alone reached accurate classification results without SWP, but this required more iterations and a larger number of features. *Fifth*, the features that were selected by SFFS originated from all available physiological sensors (EEG, EOG and ECG), indicating that data from many different sources should be fused when classifying sleepiness. This is in line with how sleep (not sleepiness) is assessed in the clinic, where polysomnography is the gold standard [65].

When iterating the SFFS algorithm 15 times, it was noted that the first few features were nearly the same across all iterations. The subsequent features that were selected, and also the number of selected features, did however fluctuate for each iteration. This indicate that as long as the few strongest features are included, the remaining features are not necessarily as important, and one might use the features that are available, or that are less costly to compute.

When testing the classifier on data from a person that it had not been trained on, the sensitivity dropped considerably, especially if SWP was not included as a feature. It has previously been shown that there are individual differences when it comes to sleepiness in general [66], as well as in driving [67], and our results further demonstrate the necessity of adapting the classifier to individual drivers. The approach chosen here was to minimize the effects of individual differences by injecting some data from that individual in the training data. Another approach would be to separately train the classifier for each individual. In the latter case, the lack of labelled training data will inevitably lead to overfitting, and the lack of labelled test data will make the test results sensitive to random noise. A practically feasible approach for large scale deployment could be to create an initial general classifier based on a dataset such as the one used here, which is then adapted to the individual driver using the large amounts of data that accumulates during everyday driving. These data could then be labelled with a biomathematical model of sleepiness, and gently incorporated in the classifier with a slow learning rate.

To the best of our knowledge, the dataset used in this paper is the largest labelled driver sleepiness dataset acquired on real roads in real traffic that is available. This gave us a unique opportunity when defining the feature sets and when optimising the classifier. Even so, this dataset is very limited. The 5 % increase in accuracy that was seen when using the full dataset for training instead of the partial training set indicates that there is a shortage of data. Splitting the dataset into three isolated parts comes at the cost of running short on data, but the alternative would be to risk data leakage from one stage in the machine learning pipeline to the next, something that would inevitably overestimate performance and cause poor generalisability to unseen data (to our knowledge, this is the only driver sleepiness classification paper that considers this aspect properly). One plausible way to increase the size of the dataset would be to augment it with data from driving simulator experiments. Just adding simulator data to the real-world dataset will however not work due to the known differences between the environments [53], [68], but by employing transfer learning in the classifier design process, this shortcoming may be overcome. On the positive side, these results indicate that machine learning for sleepiness detection has not yet reached its full potential.

The accuracy of the target values, i.e. the reliability of the sleepiness ground truth, has an impact on the design and the optimization of the classifier. Although self-reported subjective sleepiness ratings are the most commonly used ground truth [19], [34], [35], [37], [43], objective measures of sleepiness have also been considered. For example, Moller et al. [69] reported that objective measures of sleepiness and crash risk were more correlated than subjective sleepiness estimates and crash risk, suggesting a discrepancy between subjective and objective predictions of crash risks. However, van den Berg [70] found the correlation between subjective estimations and task performance to be extremely high, whereas the correlation between theta band activity and task performance was considerably lower. Similar results have been observed in a driving context. Otmani *et al.* [71] found no correlation between driving performance and power in the EEG band of 3.9–12.7 Hz (corresponding to the alpha and theta bands combined). At the same time, a significant correlation was found between driving performance and subjective sleepiness. Other ground truths of driver sleepiness that have been used for sleepiness classification include expert ratings based on video recordings [41], [46], expert ratings based on physiological signals [39], the supposed alertness level that follows from an experimental design with sleep deprived participants [47], the percentage of eye closure [38], and lane departure events. However, video based expert ratings have been found to be unreliable [72], the experimental design approach does not guarantee that the driver is alert in the supposedly alert condition, and lane departure events are rare. All in all, subjective ratings seem to be the better alternative, especially since KSS is easily applied, unobtrusive, and above all, the measure of driver sleepiness that is least affected by interindividual variations [73], [74]. A final deciding factor for using subjective ratings as the ground truth of sleepiness is that both lane position and physiological signals are used as features, and it would be unsuitable to use the same information to define the target values.

There are some limitations to this study. *First*, using physiological data as input to the classifier limits its practical usefulness, as obtrusive electrodes will never be accepted by the drivers. Instead, the developed classifier demonstrates the possibility to estimate sleepiness based on sensor data. Such a system can be used as a benchmark when developing future unobtrusive systems. It can also be seen as an intermediate step where the electrodes will gradually be replaced with wearable devices and remote sensors [75], [76]. The EOG electrodes can soon be replaced by camera-based solutions, and ECG may be measured using sensors in the steering wheel, in the seatbelt or in the seat. *Second*, the EEG electrode positions used in the three studies were based on the American Academy of Sleep Medicine (AASM) guidelines for scoring of sleep [77]. The frontal channel (Fz-A1) was chosen to pick up deep sleep waves, the occipital channel (Oz-Pz) was chosen to pick up alpha activity, and the central channel (Cz-A2) was chosen to pick up sleep spindles. However, in a sleepiness classifier, the aim is to detect sleepiness, not sleep, why other electrode configurations may be preferable. *Third*, the developed classifier does not take context nor time history into account, even though both of these factors are known to affect sleepiness [35], [78]. The presented classifier has been trained using data from rural roads and motorways, and should probably not be generalized to other environments such as urban roads. In this realm of autonomous self-driving cars, automatic sensing of the surrounding traffic environment is however becoming better by the minute. Given a large enough dataset of labelled data, there are no practical difficulties in incorporating this contextual information in the machine learning framework. Regarding time history, it could be accounted for by using hidden Markov models or recurrent neural networks as classifiers, or simply by adding time-lagged features as input to any classifier. *Fourth*, the response time of the classification system is rather slow since some of the features require up to 2 minutes of data before the results stabilize. This is of minor importance since sleepiness is a slow physiological process. *Fifth*, the developed algorithm has not been benchmarked against the current state of the art algorithms. In many cases, algorithms in the literature makes use of physiological signals that are not available in our database (EMG from the neck, respiration, 30-channel EEG etc.). In other cases, the state of the art algorithms would have to be modified before a fair comparison could be made. This involves, for example, adding a cost function to account for the imbalanced dataset, and to re-run their feature selection steps on a separate dataset to avoid data leakage. Such modifications would substantially change those algorithms, essentially turning them into new updated versions more similar to our algorithm. Under such conditions, the results of the benchmarking test would still not provide a test against the state of the art, but rather a comparison of a number of new algorithms. Such endeavours are left for future research. *Sixth*, combating driver sleepiness is not only a question of developing a reliable detection system. Once the detection is made, this information should be conveyed to the driver in a convincing way to make him/her stop driving. This, in turn, requires a full chain of countermeasures to be in place, including rest areas where it is possible to safely take a break, and education and training. The suggested sleepiness detection system is one of the many complementary pieces in this chain.

VI. Conclusions

A driver sleepiness detection system based on a random forest classifier, trained and tested with physiological data obtained from real roads in real traffic, has been developed. The results highlight the importance of personalized sleepiness detection systems. When testing the classifier on data from a person that it had not been trained on, the sensitivity dropped considerably. One way to improve the sensitivity was to add a biomathematical model of sleepiness amongst the features. Future works include taking contextual features into account, using classifiers that takes full advantage of sequential data, and to develop models that adapt to individual drivers.

Acknowledgment

This research is a collaborative effort which would not have been possible without A Anund and C Fors at VTI, T Åkerstedt and G Kecklund at the Stress Research Institute, Tuan Pham and Ingemar Fredriksson at Linköping University, and their many project partners at Volvo Car Corporation, Autoliv Inc., former Saab Automobile AB, and SmartEye AB.

References

[1] R. J. Adams *et al.*, "Sleep health of Australian adults in 2016: Results of the 2016 sleep health foundation national survey," *Sleep Health*, vol. 3, no. 1, pp. 35–42, 2017.

[2] T. A. Dingus *et al.*, "Driver crash risk factors and prevalence evaluation using naturalistic driving data," *Proc. Nat. Acad. Sci. USA*, vol. 113, no. 10, pp. 2636–2641, 2016.

[3] I. Radun, J. Radun, M. Wahde, C. N. Watling, and G. Kecklund, "Self-reported circumstances and consequences of driving while sleepy," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 32, pp. 91–100, Jul. 2015.

[4] 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.

[5] J. Connor *et al.*, "Driver sleepiness and risk of serious injury to car occupants: Population based case control study," *Brit. Med. J.*, vol. 324, no. 7346, p. 1125-1128A, 2002.

[6] J. A. Horne and L. A. Reyner, "Sleep related vehicle accidents," *Brit. Med. J.*, vol. 310, no. 6979, pp. 565–567, Mar. 1995.

[7] *Global Status Report on Road Safety*, World Health Org., Geneva, Switzerland, 2015.

[8] G. Maycock, "Sleepiness and driving: The experience of UK car drivers," (in English), *Accident Anal. Prevent.*, vol. 29, no. 4, pp. 453–462, Jul. 1997.

[9] P. Philip, F. Vervialle, P. Le Breton, J. Taillard, and J. Horne, "Fatigue, alcohol and serious road crashes in France," (in English), *Sleep*, vol. 24, p. A123, Apr. 15 2001.

[10] I. Radun and H. Summala, "Sleep-related fatal vehicle accidents: Characteristics of decisions made by multidisciplinary investigation teams," (in English), *Sleep*, vol. 27, no. 2, pp. 224–227, Mar. 2004.

[11] L. A. Reyner and J. A. Horne, "Falling asleep whilst driving: Are drivers aware of prior sleepiness?" *Int. J. Legal Med.*, vol. 111, no. 3, pp. 120–123, 1998.

[12] A. Williamson, R. Friswell, J. Olivier, and R. Grzebieta, "Are drivers aware of sleepiness and increasing crash risk while driving?" *Accident Anal. Prevent.*, vol. 70, pp. 225–234, Sep. 2014.

[13] J. M. Mullington *et al.*, "Panel discussion: Current status of measuring sleepiness," *J. Clin. Sleep Med.*, vol. 7, no. 5, pp. S22–S25, 2011.

[14] M. I. Chacon-Murguia and C. Prieto-Resendiz, "Detecting driver drowsiness: A survey of system designs and technology," *IEEE Consum. Electron. Mag.*, vol. 4, no. 4, pp. 107–119, Oct. 2015.

[15] M. Golz, D. Sommer, U. Trutschel, B. Sirois, and D. Edward, "Evaluation of fatigue monitoring technologies," *Somnologie*, vol. 14, no. 3, pp. 187–199, 2010.

[16] A. Sahayadhas, K. Sundaraj, and M. Murugappan, "Detecting driver drowsiness based on sensors: A review," *Sensors*, vol. 12, no. 12, pp. 16937–16953, Dec. 2012.

[17] A. Fernández, R. Usamentiaga, J. Carús, and R. Casado, "Driver distraction using visual-based sensors and algorithms," *Sensors*, vol. 16, no. 11, p. 1805, 2016.

[18] C. Ahlstrom *et al.*, "Fit-for-duty test for estimation of drivers' sleepiness level: Eye movements improve the sleep/wake predictor," *Transp. Res. C, Emerg. Technol.*, vol. 26, pp. 20–32, Jan. 2013.

[19] D. Sandberg, T. Akerstedt, A. Anund, G. Kecklund, and M. Wahde, "Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators," (in English), *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 97–108, Mar. 2011.

[20] D. Aeschbach, J. R. Matthews, T. T. Postolache, M. A. Jackson, H. A. Giesen, and T. A. Wehr, "Dynamics of the human EEG during prolonged wakefulness: Evidence for frequency-specific circadian and homeostatic influences," (in English), *Neurosci. Lett.*, vol. 239, nos. 2–3, pp. 121–124, Dec. 1997.

[21] C. Cajochen, D. P. Brunner, K. Krauchi, P. Graw, and A. Wirz-Justice, "Power density in theta/alpha frequencies of the waking EEG progressively increases during sustained wakefulness," (in English), *Sleep*, vol. 18, no. 10, pp. 890–894, Dec. 1995.

[22] G. Kecklund and T. Åkerstedt, "Sleepiness in long distance truck driving: An ambulatory EEG study of night driving," *Ergonomics*, vol. 36, no. 9, pp. 1007–1017, 1993.

[23] M. Simon *et al.*, "EEG alpha spindle measures as indicators of driver fatigue under real traffic conditions," *Clin. Neurophysiol.*, vol. 122, no. 6, pp. 1168–1178, 2011.

[24] J. van den Berg, G. Neely, U. Wiklund, and U. Landström, "Heart rate variability during sedentary work and sleep in normal and sleep-deprived states," *Clin. Physiol. Funct. Imag.*, vol. 25, no. 1, pp. 51–57, 2005.

[25] R. J. Apparies, T. C. Riniolo, and S. W. Porges, "A psychophysiological investigation of the effects of driving longer-combination vehicles," *Ergonomics*, vol. 41, no. 5, pp. 581–592, 1998.

[26] Y. Tran, N. Wijesuriya, M. Tarvainen, P. Karjalainen, and A. Craig, "The relationship between spectral changes in heart rate variability and fatigue," *J. Psychophysiol.*, vol. 23, no. 3, pp. 143–151, 2009.

[27] 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.

[28] G. Yang, Y. Lin, and P. Bhattacharya, "A driver fatigue recognition model based on information fusion and dynamic Bayesian network," *Inf. Sci.*, vol. 180, no. 10, pp. 1942–1954, 2010.

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

[30] N. Egelund, "Spectral analysis of heart rate variability as an indicator of driver fatigue," *Ergonomics*, vol. 25, no. 7, pp. 663–672, 1982.

[31] S. Milosevic, "Drivers' fatigue studies," *Ergonomics*, vol. 40, no. 3, pp. 381–389, 1997.

[32] R. Schleicher, N. Galley, S. Briest, and L. Galley, "Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired?" *Ergonomics*, vol. 51, no. 7, pp. 982–1010, 2008.

[33] 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.

[34] W. Kong, Z. Zhou, B. Jiang, F. Babiloni, and G. Borghini, "Assessment of driving fatigue based on intra/inter-region phase synchronization," *Neurocomputing*, vol. 219, pp. 474–482, Jan. 2017.

[35] R. Fu, H. Wang, and W. Zhao, "Dynamic driver fatigue detection using hidden Markov model in real driving condition," *Expert Syst. Appl.*, vol. 63, pp. 397–411, Nov. 2016.

[36] M. Golz, D. Sommer, M. Chen, D. Mandic, and U. Trutschel, "Feature fusion for the detection of microsleep events," *J. VLSI Signal Process. Syst. Signal Image Video Technol.*, vol. 49, no. 2, pp. 329–342, 2007.

[37] S. Hu, G. Zheng, and B. Peters, "Driver fatigue detection from electroencephalogram spectrum after electrooculography artefact removal," *IET Intell. Transp. Syst.*, vol. 7, no. 1, pp. 105–113, Mar. 2013.

[38] G. Li, B.-L. Lee, and W.-Y. Chung, "Smartwatch-based wearable EEG system for driver drowsiness detection," *IEEE Sensors J.*, vol. 15, no. 12, pp. 7169–7180, Dec. 2015.

[39] A. Picot, S. Charbonnier, and A. Caplier, "On-line detection of drowsiness using brain and visual information," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 42, no. 3, pp. 764–775, May 2012.

[40] M. Golz, D. Sommer, and J. Krajewski, "Prediction of immediately occurring microsleep events from brain electric signals," *Current Directions Biomed. Eng.*, vol. 2, no. 1, pp. 149–153, 2016.

[41] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 121–131, Jan. 2011.

[42] J. Liu, C. Zhang, and C. Zheng, "EEG-based estimation of mental fatigue by using KPCA-HMM and complexity parameters," *Biomed. Signal Process. Control*, vol. 5, no. 2, pp. 124–130, 2010.

[43] Z. Mu, J. Hu, and J. Min, "Driver fatigue detection system using electroencephalography signals based on combined entropy features," *Appl. Sci. (Switzerland)*, vol. 7, no. 2, 2017, Art. no. 150.

[44] C. Papadelis *et al.*, "Monitoring sleepiness with on-board electrophysiological recordings for preventing sleep-deprived traffic accidents," *Clin. Neurophysiol.*, vol. 118, no. 9, pp. 1906–1922, 2007.

[45] F.-C. Lin, L.-W. Ko, C.-H. Chuang, T.-P. Su, and C.-T. Lin, "Generalized EEG-based drowsiness prediction system by using a self-organizing neural fuzzy system," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 59, no. 9, pp. 2044–2055, Sep. 2012.

[46] G. Li and W.-Y. Chung, "A context-aware EEG headset system for early detection of driver drowsiness," *Sensors*, vol. 15, no. 8, pp. 20873–20893, 2015.

[47] M. Jirina, P. Bouchner, and S. Novotny, "Identification of driver's drowsiness using driving information and EEG," *Neural Netw. World*, vol. 20, no. 6, pp. 773–791, 2010.

[48] C.-T. Lin, C.-J. Cheng, B.-S. Lin, S.-H. Hung, C.-F. Chao, and I.-J. Wang, "A real-time wireless brain-computer interface system for drowsiness detection," *IEEE Trans. Biomed. Circuits Syst.*, vol. 4, no. 4, pp. 214–222, Apr. 2015.

[49] I. Damousis, I. Cester, S. Nikolaou, and D. Tzovaras, "Physiological indicators based sleep onset prediction for the avoidance of driving accidents," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Aug. 2007, pp. 6699–6704.

[50] T. A. Ranney, "Psychological fidelity: Perception of risk," in *Handbook of Driving Simulation for Engineering, Medicine, and Psychology*, D. L. Fisher, M. Rizzo, J. Caird, and J. D. Lee, Eds. Boca Raton, FL, USA: CRC Press, 2011, pp. 9.1–9.13.

[51] 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.

[52] 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.

[53] C. Fors, C. Ahlström, and A. Anund, "A comparison of driver sleepiness in the simulator and on the real road," *J. Transp., Safety Secur.*, vol. 10, no. 2, pp. 72–87, 2018.

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

[55] A. Anund, C. Fors, D. Hallvig, T. Akerstedt, and G. Kecklund, "Observer rated sleepiness and real road driving: An explorative study," *PloS ONE*, vol. 8, no. 5, p. e64782, 2013.

[56] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2352–2359, 2009.

[57] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[58] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," (in English), *Physica D*, vol. 31, no. 2, pp. 277–283, Jun. 1988.

[59] B. Jammes, H. Sharabty, and D. Esteve, "Automatic EOG analysis: A first step toward automatic drowsiness scoring during wake-sleep transitions," *Somnologie—Schlafforschung Schlafmedizin*, vol. 12, no. 3, pp. 227–232, 2008.

[60] B. Hjorth, "EEG analysis based on time domain properties," *Electroencephalogr. Clin. Neurophysiol.*, vol. 29, no. 3, pp. 306–310, 1970.

[61] W. A. Macdonald and E. R. Hoffmann, "Review of relationships between steering wheel reversal rate and driving task demand," *Hum. Factors*, vol. 22, no. 6, pp. 733–739, 1980.

[62] T. Åkerstedt, J. Connor, A. Gray, and G. Kecklund, "Predicting road crashes from a mathematical model of alertness regulation—The sleep/wake predictor," *Accident Anal. Prevent.*, vol. 40, no. 4, pp. 1480–1485, 2008.

[63] J. Mekyska *et al.*, "Assessing progress of Parkinson's disease using acoustic analysis of phonation," in *Proc. 4th Int. Work Conf. Bioinspired Intell.*, 2015, pp. 111–118.

[64] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.

[65] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," U.S. Dept. Health, Edu. Welfare, Public Health Service, Bethesda, Washington, DC, USA, Tech. Rep., 1968.

[66] N. Wijesuriya, Y. Tran, and A. Craig, "The psychophysiological determinants of fatigue," *Int. J. Psychophysiol.*, vol. 63, no. 1, pp. 77–86, 2007.

[67] 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.

[68] D. Hallvig *et al.*, "Sleepy driving on the real road and in the simulator—A comparison," *Accident; Anal. Prevent.*, vol. 50, pp. 44–50, Jan. 2013.

[69] H. J. Moller, L. Kayumov, E. L. Bulmash, J. Nhan, and C. M. Shapiro, "Simulator performance, microsleep episodes, and subjective sleepiness: Normative data using convergent methodologies to assess driver drowsiness," *J. Psychosomat. Res.*, vol. 61, no. 3, pp. 335–342, 2006.

[70] J. van den Berg, G. Neely, L. Nilsson, A. Knutsson, and U. Landström, "Electroencephalography and subjective ratings of sleep deprivation," *Sleep Med.*, vol. 6, no. 3, pp. 231–240, 2005.

[71] S. Otmani, T. Pebayle, J. Roge, and A. Muzet, "Effect of driving duration and partial sleep deprivation on subsequent alertness and performance of car drivers," *Physiol., Behav.*, vol. 84, no. 5, pp. 715–724, 2005.

[72] C. Ahlstrom, C. Fors, A. Anund, and D. Hallvig, "Video-based observer rated sleepiness versus self-reported subjective sleepiness in real road driving," *Eur. Transp. Res. Rev.*, vol. 7, no. 4, p. 38, 2015.

[73] T. Åkerstedt, D. Hallvig, A. Anund, C. Fors, J. Schwarz, and G. Kecklund, "Having to stop driving at night because of dangerous sleepiness–awareness, physiology and behaviour," (in English), *J. Sleep Res.*, vol. 22, no. 4, pp. 380–388, Aug. 2013.

[74] T. Åkerstedt, A. Anund, J. Axelsson, and G. Kecklund, "Subjective sleepiness is a sensitive indicator of insufficient sleep and impaired waking function," *J. Sleep Res.*, vol. 23, no. 3, pp. 240–252, 2014.

[75] 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.

[76] E. Bekiaris, "Monitoring sleep and sleepiness with new sensors within medical and industrial applications," *Somnologie-Schlafforschung Schlafmedizin*, vol. 11, no. 4, p. 231, 2007.

[77] R. B. Berry *et al.*, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, Version 2.2*. Darien, IL, USA: American Academy of Sleep Medicine, 2015. [Online]. Available: http://www.aasmnet.org

[78] E. Nilsson *et al.*, "Vehicle driver monitoring: Sleepiness and cognitive load," Swedish Nat. Road Transp. Res. Inst., VTI, Linköping, Sweden, Tech. Rep. 937A, 2017. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:vti:diva-11835

Henrik Mårtensson received the M.Sc. degree in biomedical engineering from Linköping University, Sweden, in 2017.

His master thesis was performed at the Swedish National Road and Transport Research Institute. He is currently a System Engineer with the Department of Human Machine Interaction, SAAB Aeronautics AB.

Oliver Keelan received the M.Sc. degree in biomedical engineering from Linköping University, Linköping, Sweden, in 2017.

He is currently a Software and Electrical Engineering Consultant, specializing in signal analysis and modeling, with ESSIQ AB, Gothenburg, Sweden.

Christer Ahlström received the M.Sc. degree in computer science and engineering and the Ph.D. degree in biomedical engineering from Linköping University, Sweden, in 2002 and 2008, respectively.

He is currently a Researcher with the Swedish National Road and Transport Research Institute, Linköping, Sweden. He specializes in biomedical engineering aspects of driver state research, including sleepiness, distraction, and workload.