Accident Analysis and Prevention 126 (2019) 95–104

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Detection and prediction of driver drowsiness using artificial neural network models

Charlotte Jacobé de Naurois[a,b,*], Christophe Bourdin[a], Anca Stratulat[b], Emmanuelle Diaz[b], Jean-Louis Vercher[a]

[a] Aix Marseille Univ, CNRS, ISM, Marseille, France
[b] Groupe PSA, Centre Technique de Vélizy, Vélizy-Villacoublay, Cedex, France

ARTICLE INFO

Keywords:
Drowsiness
Prediction
Artificial neural network
Physiological measurement
Behavioral measurement
Driving performance and activity

ABSTRACT

Not just detecting but also predicting impairment of a car driver's operational state is a challenge. This study aims to determine whether the standard sources of information used to detect drowsiness can also be used to predict when a given drowsiness level will be reached. Moreover, we explore whether adding data such as driving time and participant information improves the accuracy of detection and prediction of drowsiness. Twenty-one participants drove a car simulator for 110 min under conditions optimized to induce drowsiness. We measured physiological and behavioral indicators such as heart rate and variability, respiration rate, head and eyelid movements (blink duration, frequency and PERCLOS) and recorded driving behavior such as time-to-lane-crossing, speed, steering wheel angle, position on the lane. Different combinations of this information were tested against the real state of the driver, namely the ground truth, as defined from video recordings via the Trained Observer Rating. Two models using artificial neural networks were developed, one to detect the degree of drowsiness every minute, and the other to predict every minute the time required to reach a particular drowsiness level (moderately drowsy). The best performance in both detection and prediction is obtained with behavioral indicators and additional information. The model can detect the drowsiness level with a mean square error of 0.22 and can predict when a given drowsiness level will be reached with a mean square error of 4.18 min. This study shows that, on a controlled and very monotonous environment conducive to drowsiness in a driving simulator, the dynamics of driver impairment can be predicted.

1. Introduction

Driving a car is a complex, multifaceted and potentially risky activity requiring full mobilization of physiological and cognitive resources to maintain performance over time. Any loss of these resources can have dramatic consequences, including accidents. Moreover, the promise of autonomous vehicles makes it even more important to determine the driver's operational state. This has recently generated a large number of studies, both from the fundamental perspective and with a view to potential applications. The challenge is ambitious: not only detecting, but also predicting, degradation in the driver's operational state.

A driver's operational state while driving a car involves a complex set of psychological, physiological and physical parameters. During driving activities, several factors can be critical: in particular, fatigue and monotony may cause a loss of attention, drowsiness and even sleepiness (Dong et al., 2011). The present study focuses on a specific type of impaired operational state: drowsiness. Drowsiness is an intermediate state between alertness and sleep. In this article, we will consider drowsiness as a continuum, or scalar state. Unfortunately, drowsiness cannot be recorded directly but has to be estimated, and several estimation techniques have been proposed in the literature. These methods can be classified in different categories according to source of information: subjective assessment, sensorimotor indicators, physiological features and driving behavior and performance (Dong et al., 2011).

In the last few years, the Karolinska Sleepiness Scale (KSS), a 9-graded Lickert scale (Shahid et al., 2011), has become the most commonly employed instrument for the subjective self-assessment of drowsiness (Alhazmi, 2013; Daza et al., 2014; Friedrichs and Yang, 2010; Krajewski et al., 2009a,b; Lee et al., 2016; Li et al., 2014; Murata and Naitoh, 2015). Nonetheless, although often used, this method raises three principal issues. Firstly, the driver's state can only be assessed every 15 min, since greater frequency would probably keep the driver

* Corresponding author at: Aix Marseille Univ, CNRS, ISM, Marseille, France.
E-mail address: charlotte.jacobe-de-naurois@etu.univ-amu.fr (C. Jacobé de Naurois).

https://doi.org/10.1016/j.aap.2017.11.038
Received 27 July 2017; Received in revised form 12 October 2017; Accepted 27 November 2017
Available online 06 December 2017

awake. Secondly, according to Friedrichs and Yang (2010), when the experiment involves more than three hours of monotonous driving, the KSS becomes inadequate because drivers have difficulty judging their alertness. Lastly, subjective assessment clearly does not constitute an objective measure of drowsiness, and when the task is very monotonous, individual ratings on drowsiness differ from the person's physiological alertness level (Brown, 1997).

Features extracted from eye and head movements, classified as sensorimotor indicators, are also promising parameters to detect the operational state and are now included in many research approaches (Chen and Ji, 2012; Liu et al., 2009). Video-oculo-graphy (VOG) is commonly used to study the following features: blink frequency, blink duration and PERCLOS (PERcentage of eye CLOSure). Changes in these features are considered under low-level control, offering an easy way to monitor the activity of the neurovegetative system (Caffier et al., 2003; Wierwille and Ellsworth, 1994). These features are generally extracted with image processing algorithms based on eye, head and gaze movement tracking. Thus, the quality of the estimation is highly dependent on this first signal-processing step.

Physiological features are also frequently used to assess drowsiness because they are continuously available and could be considered as an objective, more direct, measure of the functional state. The main recordings of signals related to drowsiness are the electroencephalogram (EEG), the electrocardiogram (EKG) and electro-dermal activity (EDA) (Borghini et al., 2014; Dong et al., 2011). The gold standard appears to be the EEG, the most direct indicator of central nervous system activity (De Gennaro et al., 2001). However, the EEG is quite intrusive, and proper installation of an extensive set of electrodes on the participant's scalp requires expertise and time. It has been established that when a change in vigilance is observed, changes on psychophysiological arousal can be also observed, and these changes can be monitored by measures of the central and autonomic nervous system activity (Haarmann and Boucsein, 2008). Concerning EKG, since heart rate variability (HRV) is linked to the autonomic nervous system this feature is often used as an indicator of drowsiness because change on HRV can provide information about the autonomic nervous system (Elsenbruch et al., 1999; Lal and Craig, 2001; Riemersma et al., 1977; Stein and Pu, 2012). Moreover, some studies on drowsiness, vigilance or workload also record and analyze respiration rate and amplitude (Besson et al., 2013; Ju et al., 2015; Reimer et al., 2009; Rodriguez Ibañez et al., 2011).

Yet a direct relationship between physiological features and cognitive state is hard to define, because these physiological features vary with other states (including, but not limited to, emotion, workload, physical fatigue) or with the context. These variations according to state also differ from one person to another. Thus, each physiological indicator has its own limits. Heart rate usually decreases during driving and when the driver is tired (Lal and Craig, 2001), but the opposite may also occur (Apparies et al., 1998). Peiris et al. (2005) showed that two independent experts analyzing EEGs to detect drowsiness may not make the same assessment for the same participant at the same time. On the other hand, EDA can be influenced by stress (Healey and Picard, 2005) and emotions (Rebolledo-Mendez et al., 2014). Taken alone, therefore, these indicators in themselves cannot be considered as adequate and exclusive indicators of drowsiness or fatigue.

Driving behavior and performance analyses have the main advantage of being non-intrusive. Some signals such as pressure on pedals or car movements are easily available. The standard deviation of car position relative to lane midline (also named standard deviation of lane position (SDLP)), and steering wheel movements, are the most common features used to detect drowsiness (Arnedt et al., 2001; De Valck et al., 2003; Liu et al., 2009; Philip et al., 2004). However, here again, driving performance and activity are not specific indicators of drowsiness. For example, driving performance can decrease with other factors such as distraction (Tango et al., 2009), or with a decline in attention (Marin-Lamellet et al., 2003)

Since none of these feature families is consensually considered as a specific indicator of drowsiness, various measures are often used jointly. Such a hybrid approach minimizes the number of false alarms while maintaining a high rate of recognition (essential for good acceptance of the system by the human operator, Dong et al., 2011), mainly because no signal emerges as the reference marker allowing real time measurement that is both relatively non-invasive and reliable. Moreover, there is no direct link between all these features and the "operational state", which is why methods such as machine learning or statistical models are used, combining the different measures.

The different algorithms used include k-nearest neighbors (Chauhan et al., 2015), decision trees (Lee et al., 2010; Sukanesh and Vijayprasath, 2013), Bayesian classifiers (Lee and Chung, 2012; Yang et al., 2010), Support Vector Machines (Bhowmick and Chidanand Kumar, 2009; Krajewski et al., 2009a,b; Liang et al., 2007; Yeo et al., 2009), artificial neural networks (ANN) (Bundele and Banerjee, 2009; Eskandarian et al., 2007; Sayed and Eskandarian, 2001; Samiee et al., 2014), ensemble methods like random forest (Krajewski et al., 2009a,b; McDonald et al., 2013; Torkkola et al., 2008; Zhang et al., 2004) and, more recently, deep learning (Hajinoroozi et al., 2015). Most studies consider the problem of estimating the driver's impaired operational state as a classification problem. Is the driver in an impaired state or not? Is the driver drowsy or not? However, the evolution of the state of the driver can also be considered as a regression problem, i.e. the driver goes through various continuous states, although regression models are rarely used in the literature (Murata and Naitoh, 2015). Nonlinear modeling machine learning (such as with ANNs) is also often used. With these techniques, the model can extract information from noisy data, and can avoid over-fitting, making it generally more robust (Dong et al., 2011). Since in the context of driving we expected over-fitting and noisy data, the present study uses machine-learning techniques based on artificial neural networks.

Most research focuses on the detection/estimation of an impaired state, rather than on its prediction, even though they adopt the term "prediction" (Chen, 2013; Hargutt and Kruger, 2001; Ji et al., 2004; Verwey and Zaidel, 2000). This is because in machine learning, the term "prediction" is used to infer the label of an object not seen during the learning phase. However, some studies try to predict what the ground truth will be in the subsequent few minutes: the ground truth was shifted for one epoch (Kaida et al., 2007), while different lags (+1, +2, +3, +4, +5, +7, +10 min) were tested by (Larue, 2010). Murata et al. (2016) obtained the highest prediction accuracy using the data between 20 and 120 s before the prediction. Watson and Zhou (2016) detect micro-sleep with 96% accuracy and are able to predict, between 15 s and 5 min in advance, the time when the next micro-sleep will occur. However, the time when the first micro-sleep occurs obviously cannot be predicted by such methods.

As explained above, using a single source of information does not seem to be an efficient way to accurately assess the state of the driver. Different sources of information and different models are used in the literature, and results are hard to generalize away from well-controlled laboratory conditions. In the present study, we collected information originating from different sources: physiological, behavioral, and psychological data from the driver, as well as performance information from the vehicle. The goal of this study is to develop and evaluate a model with an artificial neural network (ANN), so as to predict when a given impaired state will be reached in addition to detecting this impaired state. We deliberately chose unobtrusive recording techniques easily applicable in a car. Different datasets using different sources of information were tested, to determine which kind of information yields the most powerful model. We put forward two hypotheses. First, we hypothesized that it is possible to predict when the impaired state will arise by using the sensorimotor, physiological and performance indicators used to detect drowsiness. Second, we hypothesized that adding information such as driving time and participant information will improve the accuracy of the model.

2. Materials and methods

2.1. Participants

A total of 21 participants were included in the study (mean age ± SD: 24.09 ± 3.41 years; 11 men and 10 women). On the day of the experiment, the participants were not allowed to drink alcohol, coffee or tea. Inclusion criteria were: valid driver's licence for at least 6 months, no visual correction needed to drive, not susceptible to simulator sickness (as assessed by the Motion Sickness Susceptibility Questionnaire, Short-form (MSSQ-Short, Golding, 1998) and an Epworth scale score (assessing susceptibility to drowsiness) below 14 (Johns, 1991). A score of below 8 on this scale means the person has no sleep debt. A score of from 9 to 14 means the person shows signs of sleepiness, and if the score is above 15, the person shows signs of excessive sleepiness. Before the experiment, participants were questioned on their age, their quality of sleep (on a scale of 1–10), their caffeine consumption (never, rarely, one or two cups per day, more than two cups per day), driving frequency (occasionally, several times a month, a week or a day), number of kilometers per year. To assess their circadian typology, their score on the Horne and Ostberg morning/evening questionnaire (Horne and Ostberg, 1975) was also noted. All these indicators concerning the participants were later considered as participant information, and used with a view to improving the performance of the model.

2.2. Protocol

The participants drove during between 100 and 110 min in a static driving simulator in an air-conditioned room with temperature control set at 24° Celsius, after lunchtime. According to the literature about circadian rhythms, the probability of falling asleep between 02:00 to 06:00 and 14:00 to 16:00 is 3 times higher than at 10:00 or at 19:00, respectively (Horne and Reyner, 1999). We chose a period corresponding to an intermediate level between a low risk of drowsiness (in the morning) and the highest risk (end of the night). The road and traffic were generated with SCANeR Studio®. While driving, data on driving performance, eyelid and head movements, and physiological data were recorded using the following hardware and software: SCANeR Studio® for driving performance at 10 Hz, faceLAB® for sensorimotor signals at 60 Hz, and EKG, pulse plethysmography (PPG), EDA and Respiration with the Biopac® MP150 system and Acqknowledge® software at 1000 Hz. In this study, EDA was also recorded but not used due to extensive signal loss. A webcam was placed on top of the central screen of the simulator to video-record the participants during the session.

At the beginning of the session, the participants drove along a highway for roughly 90 min, then turned off the highway and drove for around 5 min to reach a city. Finally, they drove in an urban environment for roughly 5 min. During most of the highway stretch, there was no traffic. Some 2/3 of the way along, 22 cars appeared from the right of the highway, disappearing a few kilometers later. This sudden addition of traffic was intended to change the driver's level of drowsiness. Rossi et al. (2011) demonstrated that a driver is more susceptible to sleepiness in a simulator with a monotonous scenario, and during the afternoon.

2.3. Data analysis and modeling

The level of drowsiness, the so-called ground truth (indeed, the real state of the driver is not directly accessible and must be evaluated), determined as a reference in this study is based on subjective assessment by video analysis and independently coded by two raters. Their evaluation was based on a method proposed by Wierwille and Ellsworth (1994), which used a scale between 0 and 100. For practical reasons in relation with the ANN, we decided to use a smaller scale (from 0 to 4 with a step of 0.5) as proposed by Belz et al. (2001). This ground-truth determination method was chosen because the assessment by video coding is reliable and allows a comprehensive assessment of the driver state. Other methods, such as questionnaires (e.g. KSS), reaction time to a double task or even EEG are quite invasive and may disturb the driver and thus influence his/her state. However, video analysis is long and requires several observers with a certain level of training. In order to be more reliable, this method can use criteria and rating scale as a basis for different observers. The ORD relies on a continuous scale from "alert" to "extremely drowsy" with a list of criteria which can be observable in the driver, characteristics of a drowsy driver (Wierwille and Ellsworth, 1994).The two trained raters evaluated each minute of video and rated each segment on a scale ranging from 0 (alert) to 4 (extremely drowsy). The mean of the two raters was taken as the drowsiness level. Inter-rater reliability was computed with the Pearson's linear correlation ($R = 0.71$ and $p = 0.00$).

In order to synchronize data obtained at various sampling frequencies, we averaged data over periods of 1 min. Thus, the final sampling rate is 1/min for each feature, including ground truth.

The modeling process can be divided into two phases. First, one Artificial Neural Network (ANN) detects the level of drowsiness from a predetermined set of features (detection model). This ANN is used to detect the impaired state (level of drowsiness). Second, if drowsiness is under 1.5, a second ANN predicts (in min) when it will reach 1.5 and gives this time as its output (for instance when the level is reached), otherwise its output is 0 (prediction model). The threshold was set at 1.5 for the following reason. McDonald et al. (2013) defined the limit between "not drowsy" and "drowsy" at a level between 1 and 2 (0 or 1, not drowsy; 2, 3, 4: moderately, very or extremely drowsy). We chose the level of 1.5 as a threshold for defining the impaired state because this level means that at a given time, one of the two raters has evaluated the state of the participant as moderately drowsy (level 2) while the other evaluated the state as 1. These two ANNs were trained independently.

The neural network toolbox (Beale et al., 1992) of Matlab R2013a was used to create the ANNs. Two feedforward neural networks were used with 2 hidden layers, and a back propagation training method was applied using the Levenberg-Marquardt algorithm (Levenberg, 1944). The error was validated by ten-fold cross-validation and a search grid. The performance function used for learning was the mean squared error (the average squared error between the network outputs and the target output). To avoid overfitting, the total dataset was distributed in a training sub-dataset (70% of the total set, to learn the network's node weights), a validation sub-set (15%: to stop learning and avoid overtraining) and a testing sub-set (15%: to evaluate the model's ability to work on previously unseen data. This property is also called 'generalization').

In addition, three other metrics were used to evaluate the model: first, the percentage of numbers of absolute errors below a threshold (0.5 for detection of degree of impairment and 5 min for predictions and for the testing dataset: the higher this metric, the better the model performs); second, the range of errors containing 95% of the values; and third, the coefficient R of the correlation between outputs and targets.

Driving performance and driving behavior indicators (car dataset) used in the model were: lateral distance relative to the midline, time-to-line-crossing (Bergasa et al., 2006), steering wheel angle, accelerator pedal angle, shift relative to the lateral line, speed, and number of line crossings. Physiological features used in the model (physiological dataset) were the heart rate and its variability, and the respiration rate and its variability. Sensorimotor features (behavioral dataset) extracted from FaceLab data were blink duration and its frequency, PERCLOS, head movement in translation and rotation, and saccade frequency. Participant information recorded consisted of score on circadian typology, score on Epworth scale, sleep quality, driving frequency, number of cups of coffee a day and age. Driving time (the time elapsed

Table 1
All the variables (grouped by source of information, in column) computed for each participant for each minute, used as input for ANNs.

Physiological measurements	Behavioral measurements	Car measurements
HR: Heart Rate (average and standard deviation) (beat/min)	Blink duration (average and standard deviation)	Lateral distance from the closest lane and the center of the car in m (average and standard deviation)
S_{vlf}: HR signal Very Low Frequency Power (0.0-0.04 Hz)	Blink frequency (average and standard deviation) (per minute)	Time to lane crossing (average and standard deviation)
S_{lf}: HR signal Low Frequency Power (0.04-0.15 Hz)	PERCLOS (average and standard deviation) (% of eye-closure time)	Steering angle (average and standard deviation)
S_{hf}: HR signal High Frequency Power (0.15-0.4 Hz)	Head position x (average and standard deviation)	Steering angle velocity (average and standard deviation)
S_{vhf}: HR signal Very High Frequency Power (0.4-3.0 Hz)	Head position y (mean and standard deviation)	Steering entropy (computed from steering angle)
Sympathetic ratio ($S_{lf}/(S_{vlf} + S_{lf} + S_{hf})$	Head position z (average and standard deviation)	Number of direction change (0-crossings) per minute (computed from steering angle)
Vagal ratio ($S_{hf}/(S_{vlf} + S_{lf} + S_{hf})$	Head rotation x (average and standard deviation)	Accelerator pedal angle (average and standard deviation)
Sympathetic-vagal ratio (S_{lf}/S_{hf})	Head rotation y (average and standard deviation)	Lateral shift of the vehicle center relative to the lane center (average and standard deviation)
Respiration Rate (average and standard deviation) (per minute)	Head rotation z (average and standard deviation)	Vehicle speed (km/h) (average and standard deviation)
	Saccade frequency (mean and standard deviation) (per minute)	Number of out-the road per minute

since the beginning of the driving session, in minutes) was also used as an input feature for the model (see Table 1). In an attempt to rebase individual differences, we subtracted from each signal the mean of the first five minutes of this signal, so that the signal represents variation from an initial state. To optimize learning, each feature was normalized such that minimum and maximum values lie within [−1;1].

3. Results

The ANNs were trained 16 times (4 × 2 × 2) with different datasets. Each dataset results from the combination of the following: the three sources of information tested alone or all together (thus 4 combinations), with or without elapsed time (2 cases) and with or without information about the participants (2 cases). The Tables 2 and 3 present the performance obtained with each of the 16 datasets. In this section, the results will be presented with the driving time (labeled with '1′ in tables) and without (labeled with '0′ in Tables 2 and 3), with the information about the participant (labeled with '1′ in tables) and without (labeled with '0′ in Tables 2 and 3). The grouping was decided according to how these variables were recorded in our experiment (and possibly in a real car), that is to say with which equipment. Indeed, the vehicle information can be recorded from the vehicle's Controller Area network (in our experiment with SCANeR® software), the behavioral measurements with a camera and a specific image processing system (in our experiment with faceLAB®) and physiological measurements with physiological sensors and an A/D system (in our experiment with Biopac®, in a real car it could be with a smart-watch).

3.1. Detection

In this section, we present model performance in detecting drowsiness level, as defined by the ORD scale (from 0 to 4, see Methods section). The error is the difference between the real state (as given by the subjective evaluation, the so-called ground truth) and the output, squared and averaged over epochs to provide the mean squared error of the trained model.

From an absolute point of view, the dataset configuration providing the best performance (lowest mean square error) in training the model contains driving time, participant information and behavioral features (# in Table 2). With this dataset, the mean square error is 0.22 ± 0.02 and more than 80% of the absolute value of the error of the testing data is under 0.5 (less than one-half of a state level, as defined by the ORD scale). Ninety-five percent of the absolute value of the error is under 0.87. In other words, the model is off by less than one drowsiness level on our scale, in 95% of cases. Performance is similar when car information is included. The mean square error is 0.23 ± 0.06. More than 86.34% of the absolute value of the error of the testing data is under 0.5. Ninety-five percent of the absolute value of the error is under 0.73, i.e. in 95% of cases the model is off by less than one drowsiness level on our scale.

Table 2
Model performance in detecting drowsiness level for the testing dataset: mean square error (MSE), standard deviation (STD), according to dataset used, with (1) or without (0) driving time, with (1) or without (0) participant information. The worst performance (highest MSE) is highlighted in bold and with a * while the best performance (lowest MSE) is highlighted in bold and with a #.

Driving Time	Participant information	Dataset	Source	MSE	STD	|Error |95%	% Error < 0.5
0	0	Testing	All	0.43	0.04	1.16	0.63
0	0	Testing	Behavioral	0.42	0.02	1.16	0.64
0	0	Testing	Car	0.69	0.04	1.48	0.50
0	0	Testing	Physiological	**0.81***	0.05	1.51	0.43
0	1	Testing	All	0.41	0.04	1.10	0.62
0	1	Testing	Behavioral	0.39	0.04	1.14	0.69
0	1	Testing	Car	0.62	0.03	1.34	0.54
0	1	Testing	Physiological	0.76	0.03	1.52	0.44
1	0	Testing	All	0.27	0.02	0.91	0.80
1	0	Testing	Behavioral	0.23	0.02	0.80	0.83
1	0	Testing	Car	0.40	0.05	1.20	0.66
1	0	Testing	Physiological	0.38	0.05	1.06	0.70
1	1	Testing	All	0.24	0.02	0.84	0.81
1	1	Testing	Behavioral	**0.22#**	0.02	0.87	0.80
1	1	Testing	Car	0.23	0.06	0.73	0.86
1	1	Testing	Physiological	0.29	0.07	0.75	0.82

Table 3
Performance of the model in predicting drowsiness level with the testing dataset: mean square error (MSE), standard deviation (STD), according to whether dataset is used with (1) or without driving time (0), participant information, and source of recorded information. The * symbol indicates the worst performance and the # symbol the best performance. The best and worst performance are also higlighted in bold.

Driving Time	Participant information	Dataset	Source	MSE	STD	|Error |95%	% Error < 5
0	0	Testing	All	33.64	7.63	9.29	0.79
0	0	Testing	Behavioral	23.61	3.15	8.12	0.86
0	0	Testing	Car	**60.09***	6.19	13.12	0.73
0	0	Testing	Physiological	43.77	6.24	11.47	0.74
0	1	Testing	All	28.26	2.82	8.79	0.82
0	1	Testing	Behavioral	22.83	4.03	7.98	0.89
0	1	Testing	Car	50.22	8.84	12.11	0.73
0	1	Testing	Physiological	41.82	4.11	11.83	0.74
1	0	Testing	All	10.64	3.39	4.26	0.97
1	0	Testing	Behavioral	5.46	1.50	2.92	0.99
1	0	Testing	Car	31.14	10.73	6.25	0.93
1	0	Testing	Physiological	15.97	1.70	7.01	0.89
1	1	Testing	All	7.69	2.17	3.12	0.98
1	1	Testing	Behavioral	**4.18#**	1.17	1.98	0.99
1	1	Testing	Car	4.67	1.33	2.43	0.99
1	1	Testing	Physiological	5.51	1.84	2.62	0.98

When neither driving time nor participant information is used (line 0-0 in Table 2), or when only one of these is used (0-1 or 1-0), the model performs better with all datasets used together or with the behavioral dataset used alone; performance is slightly worse with the physiological or car datasets used alone. As stated above, the model performs best, for each dataset or for all three datasets used together, when both driving time and participant information are included (1-1).

Figs. 1 and 2 present, respectively with (Fig. 1) and without (Fig. 2) driving time and participant information, the frequency histogram of distribution of error (left panel, A) and the correlation (right panel, B) between real state (target, horizontal axis) and estimated state, the output of the ANN (vertical axis). The model is trained with behavioral data in Fig. 1 and with all datasets in Fig. 2; thus, Fig. 1 illustrates the best, and Fig. 2 the worst, performance for the training, validation and testing datasets. Linear regressions were applied to the output of the model to correlate them with the ground truth. With a perfect model, all data points would be on the diagonal line of the correlation graph. Fig. 1 shows that, for each of the three datasets, simulated values are well correlated with expected values (ground truth). The R-values are actually very close to unity (0.93, 0.91, 0.91 respectively for the training, validation and testing datasets). Moreover, the slopes of the regression lines are very close to unity (0.87, 0.88, 0.88 respectively for the training, validation and testing datasets) and the intercepts are close to zero (0.17 for all three datasets). Errors are calculated, at each 1 min epoch, as the difference between the output of the model and the ground truth. The graph on the left of Fig. 1 shows a peak at 0.05, meaning that most of the errors are close to 0. Also, more than 95% of the instances had an error of between −1.16 and 1.16. In Fig. 2, the correlations between output and target are still good but there is greater variability (R = 0.87, 0.74, 0.78 respectively for the training, validation and testing datasets). The model used for the results presented in Fig. 2 (behavior, physiology and car) is less accurate than the model which results are presented on Fig. 1 (behavior, elapsed time and participant information). As for errors, the graph on the left shows a single but broader peak at 0.2 and −0.02, also meaning that most of the errors are close to 0.

3.2. Prediction

This section presents the performance of the second model, aimed at predicting when a driver will reach a given drowsiness level (here 1.5). The error, for each epoch, is the difference between the time remaining

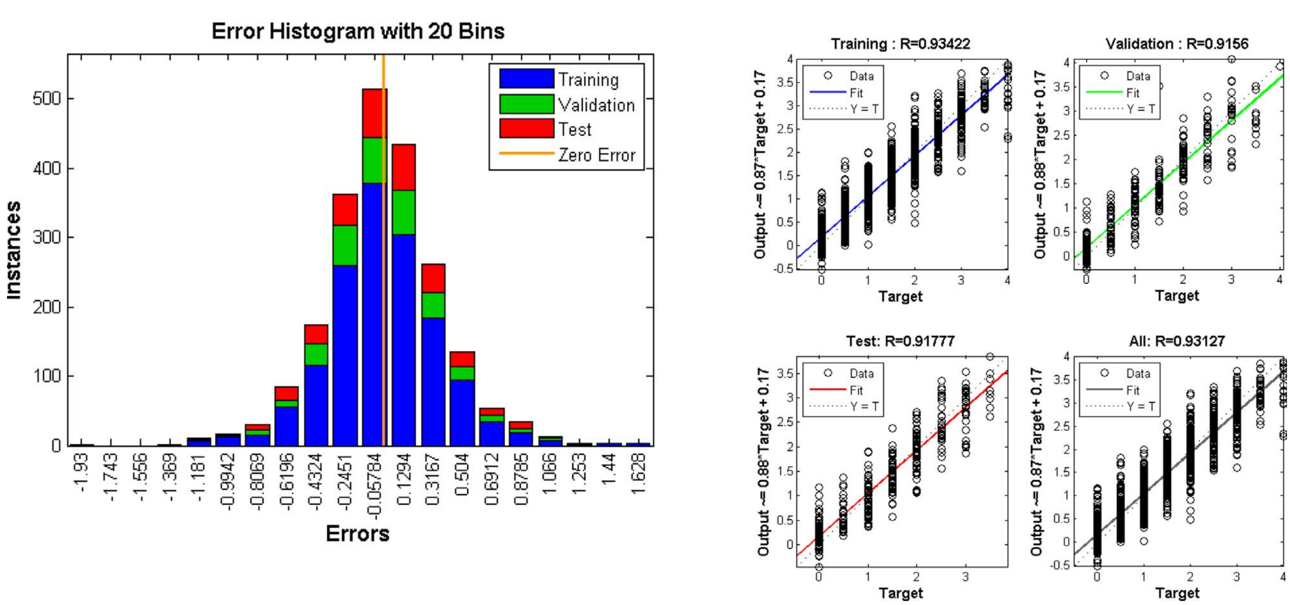

Fig. 1. frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated state, for a model trained with behavioral dataset, driving time and participant information.

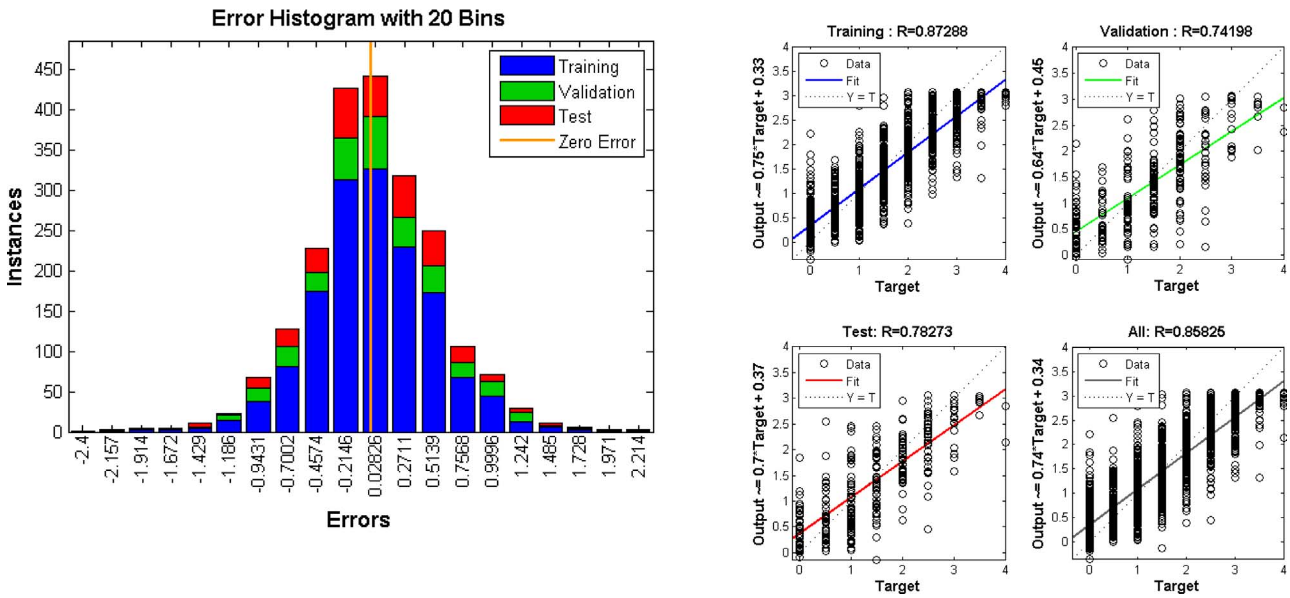

Fig. 2. frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated state, for a model trained with behavioral, car and physiological datasets.

from the current epoch before the target level is really reached (as per the subjective evaluation) and the time predicted by the trained model (squared and averaged over epochs to provide the mean squared error).

The best performance is achieved with a combination of driving time, participant information and the behavioral dataset. The mean square error is 4.18 ± 1.17 min. For 95% of the testing data, the absolute value of error is under 2 min and more than 99% of the absolute value of error is under 5 min. Similar, but not higher, accuracy is achieved with the car and physiological datasets (4.67 ± 1.33 and 5.51 ± 1.84). Ninety-five percent of the absolute value of error is under 2.43 and 2.62, respectively. For more than 97% of the testing data, the absolute value of error is under 5 min.

The worst model performance in predicting drowsiness is with the car dataset alone (60.09 ± 6.19 min). Performance improves with the addition of participant information (50.21 ± 8.84 min), or of driving time (31.14 ± 10.72 min). The model becomes very accurate when both driving time and participant information are included with the car dataset (4.67 ± 1.33 min).

For each source of information (all, behavioral, car and physiological datasets), the model is more accurate when both driving time and participant information are included in the dataset than with either driving time or participant information alone, or with no additional information.

Figs. 3 and 4 present the frequency histogram of distribution of errors (left panel) and the correlation between real time (target, horizontal axis) and estimated time (vertical axis) of appearance of drowsiness, respectively with (Fig. 3) and without (Fig. 4) driving time and participant information, The model is trained with behavioral data in Fig. 3 and with all datasets in Fig. 4, so that Fig. 3 illustrates the best, and Fig. 4 the worst, performance. On Fig. 3, the graph on the right shows that the relation between target and output is very precise, data are close to the diagonal (very high R, better than 0.98 for the training, validation and testing datasets, the slopes are better than 0.99). On the left part of Fig. 3, the main peak is at 0.3, meaning that the model has an error inferior at 0.3.

4. Discussion

Detecting impairment of a driver's operational state is a major safety issue, addressed in numerous studies. While recent car models go some way towards providing this detection capacity, it is clear that recent technological developments are not sufficient to meet the challenge of safety in modern vehicles. Predicting the degree of driver impairment, and when it will occur, remain important research objectives requiring more complex treatment of heterogeneous information from diverse sources. The objective of this study was to assess whether the time of occurrence of a given state of drowsiness could be predicted by using ANN models (one to detect drowsiness and a second one to predict drowsiness).

Overall, our results demonstrate that, using an ANN trained with the same information used to detect drowsiness, it is possible to predict when a driver's impairment will appear to an accuracy of approximately 5 min. Moreover, to further improve accuracy, external information such as driving time or a driver profile can be added to the model. In his study, Larue (2010) accurately predicted a driver's decreased vigilance up to five minutes in advance, and up to 10 min in advance with 70% to 80% accuracy. Under quite different conditions, and with different types of information, our model seems to be more accurate. In our worst case, for 95% of the test dataset, the model can predict when the impairment will appear to within 13.11 min. In our best case, for 95% of the test dataset, the model can predict the impairment to within 1.97 min.

As explained in the results section, model performance, both on detecting a drowsiness level and on predicting when this level will be reached, varies considerably according to the datasets used to train the model. This raises the question of the relevance of using physiological signals, behavioral features and driving activity, and of the respective roles of these different datasets in model performance. An important point highlighted by our results is how temporal (driving time) and idiosyncratic (participant information) data impact model performance. The limitations of our model with regard to generalization (i.e. the ability of the model to accurately treat previously unseen data), and from a more general point of view, inter-individual variability, will also be discussed.

4.1. Dataset comparison: behavioral/physiological/car

Our objective was to use the same information both to detect drowsiness and to predict the time when a given drowsiness level would be reached. Interestingly, when trained with all datasets, either singly or in combination, the model gave satisfactory results. The dataset giving the best performance is the behavioral dataset (followed by the

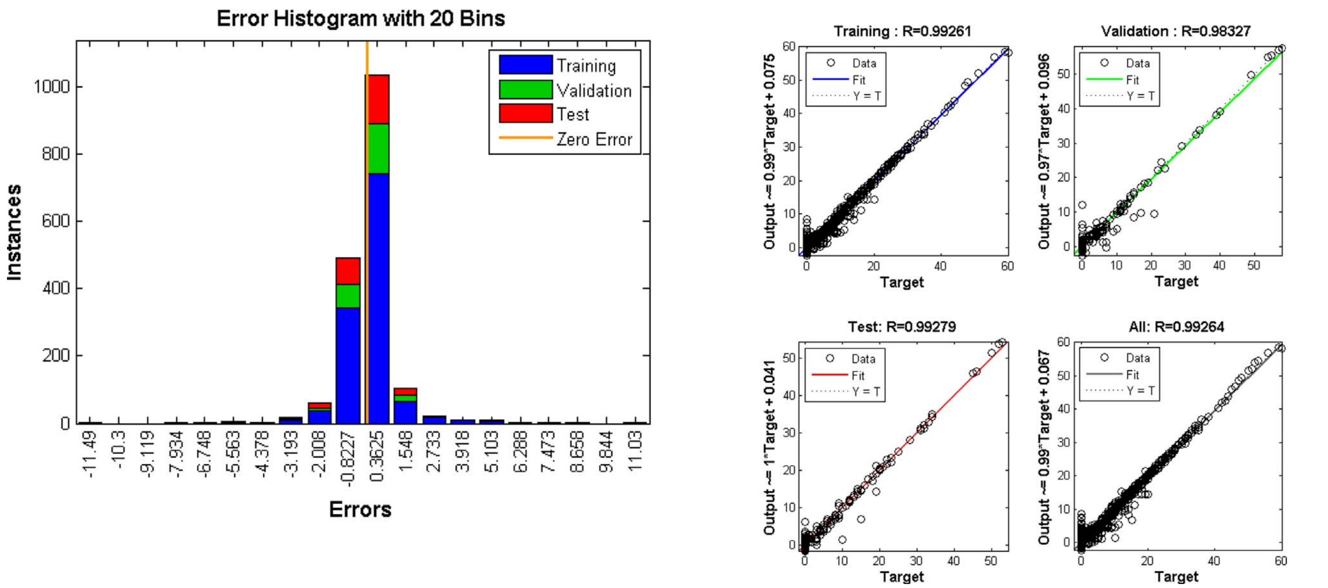

Fig. 3. frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated times, for a model trained with behavioral dataset, driving time and participant information.

physiological dataset and finally the car dataset), both in detecting the degree of drowsiness and in predicting when a given drowsiness level will occur. Similar results were previously reported. Samiee et al. (2014) showed that information about blinks leads to highly accurate detection (90.74% detection of a drowsy state), while lateral deviation of the car and steering wheel angle provide 85.37% and 87.22% accuracy, respectively. However, when all three sources of information (blinking, lateral position and steering angle) were used together, accuracy increased to 94.69%, although this was not borne out by our study. As in our study, Daza et al. (2014) obtained better results with features extracted from eyelid movement (such as PERCLOS) than with features extracted from driving behavior. In the literature, HRV data showed a correlation with drowsiness (Elsenbruch et al., 1999; Lal and Craig, 2001; Stein and Pu, 2012). Yet our model gave better results with ocular and head parameters than with physiological variables: the ORD scale showed a stronger correlation with the ocular parameters than with physiological variables such as EKG and Respiration (Rost et al., 2015). Wang and Xu (2016) consider eye features as the prime input for detection of drowsiness. However, since they are usually computed by image processing, these features cannot be considered fully reliable. Although techniques have progressed considerably in recent years, detecting face and gaze movements remains tricky in complex situations (for example, subjects with glasses, variable or low light conditions, Benoit and Caplier, 2005; Friedrichs and Yang, 2010).

Our behavioral, physiological and, to a lesser extent, car datasets led to the best model performance. With all sources of information in the same neural network, performance could be expected to improve because the neural network can better learn dependencies between different kinds of information. Unfortunately, our results do not bear this out. A single ANN-based model may not be the best way to take advantage of the dependencies between the different sources of information. An alternative, inspired by Samiee et al. (2014), might be to linearly combine the outputs of three ANNs, each trained with a different dataset: car, physiological or behavioral.

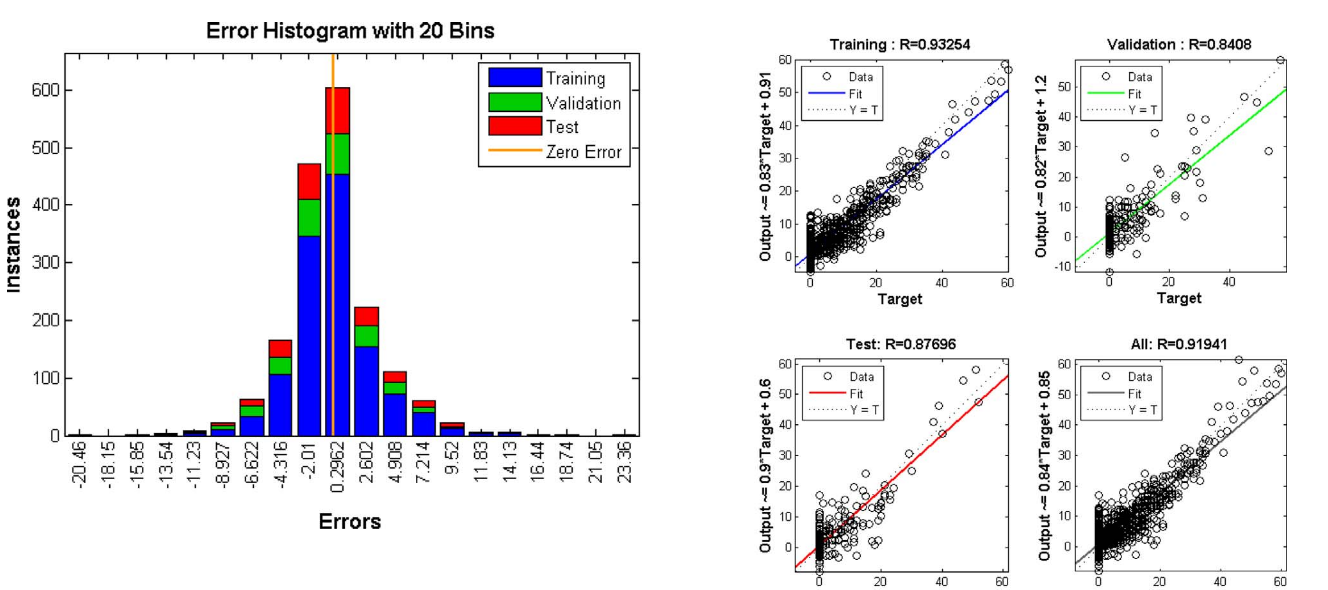

Fig. 4. frequency histogram of error distribution (left panel) and correlation (right panel) between real and estimated times, for a model trained with behavioral, car, and physiological datasets.

Surprisingly, other information often included in the literature appears less relevant here. For instance, car deviations relative to the road and line crossings are often considered as signs of drowsiness (Philip et al., 2005). Yet our results unexpectedly show that a model trained with the car dataset alone is less accurate than models trained with other datasets. This may be due to the fact that driving activity and performance are non-linearly correlated with degree of drowsiness. Thus, they may be more useful to detect a critical state (very or extremely drowsy) than to assess a monotonous evolution of the driver's state (alert, slightly, moderately, very, extremely drowsy). Ingre et al. (2006) showed that the SDLP score (Standard Deviation of the Lateral Position) dramatically increased with a subjective measure of drowsiness (KSS scale: from 1 to 9). Since our postulate was to consider drowsiness as a continuous variable, the car dataset was obviously not the most appropriate for training our model.

Finally, a potential bias in detection might be suspected from the fact that the ground truth is based on subjective evaluations from video recordings of the participant's motor behavior, which could be thought to explain the superior performance of the behavioral dataset. However, it is worth noting that these features are consensually described in the literature as the most objective and pertinent indicators of drowsiness. It is therefore difficult to conclude on whether the high performance of a model trained with behavioral data is due to the way ground truth is set or to the greater relevance of this particular set of data.

4.2. The role of driving time

Driving time (the time elapsed since the beginning of the driving session, in minutes) plays an important role here, greatly improving the performance of the model. Obviously, the longer a driver drives under monotonous conditions, the greater the probability of being drowsy (Philip et al., 1999a,b). This is why drivers travelling on highways are often reminded to take a rest break after two hours of driving (Philip et al., 1999a,b). Thus, the model can be considered to have learned a linear relationship between elapsed time and the remaining time before the occurrence of the critical state (naturally until the critical level is reached, after that the predicted time will be 0). It could therefore be deduced that driving time is sufficient per se to predict impairment of the driver's state. However, our experiment showed that participants reached a critical level at different times after the session began. Some participants reached the critical state as early as 10 min after the beginning of the driving session, and others after around 30 min. Moreover, we observed that some participants could be drowsy at a particular time and subsequently become alert again. It can therefore be concluded that there is not a simple linear relationship between driving time and the time before a given drowsiness level is reached. To determine the real weight of driving time, we consecutively trained two models with this sole feature, and then tested their detection (model 1) and prediction (model 2) capabilities. For the detection of the drowsiness level, the mean square error was 0.47 ± 0.54. For the prediction of the time before the drowsiness level is reached, the mean square error in the generalization phase was 17.77 ± 2.15 min. Interestingly, we find that the models trained with driving time alone perform better than models trained with car or physiological datasets alone, but worse than models trained with behavioral dataset alone or with behavioral, car and physiological datasets combined. This shows that, while driving time is a good predictor of drowsiness, it is not the best.

Secondly, a model based on driving time alone would be unable to account for wakening events, such as a rest period or a traffic change. For instance, caffeine is reported to reverse time-on-task degradation of performance on sleep-deprived participants (Wesensten et al., 2004). A short nap or rest may counteract drowsiness (Anund et al., 2015). Thus, if the driver drinks a cup of coffee or takes a rest, a model based on driving time alone would need to be reinitialized. How and when this reset should be performed is an important question, requiring further experiments.

4.3. Generalization and inter-individual variability

Generalization is highly relevant in an industrial context. However, we cannot prove that our model can be generalized to new participants whose data have never been used to train the model. Inter-individual variability (sensitivity to drowsiness, behavioral, physiological or psychological idiosyncrasies) may be a limiting factor for generalization (how the model behaves with previously unseen data) and transfer (how knowledge acquired in a given domain can be adapted to another domain). In our study, the data subset used for the tests (e.g. to evaluate the model's ability to treat previously unseen data, also called the 'generalization process') was randomly chosen among the full set of data from all subjects. Thus, at this stage, it is not possible to determine whether the algorithm would perform well with the full dataset for a given subject whose data were not used to train the model. To do so would require multiple replications of the experiment under the same conditions over a longer period.

It is a major challenge to find a general model which can be trained with a limited number of drivers and then applied to other drivers (Karrer et al., 2004), due to inter-individual variability. Many studies (for a review see Liu et al., 2009) reported great variability in how drowsiness affects performance and physiological parameters in general. It is now recognized that neurobehavioral and cognitive performances vary considerably from one individual to another (Van Dongen et al., 2004a, 2004b). For instance, Philip et al. (2004) studied cognitive performance after sleep deprivation. They found that performance was highly impaired, but more so in elderly participants than in younger participants. In car driving, according to Ingre et al. (2006), there is extensive inter-individual variability in driving behavior and eye behavior: under similar conditions, individuals can present differing profiles of drowsiness evolution over time, and for a given self-declared drowsiness level, markers such as eye blink duration also vary considerably. In our study, participant information (like age or circadian activity) significantly improved accuracy both in detection and in prediction. These results point in the same direction as those of Wang and Xu (2016), who found that including individual factors improved accuracy. Sensitivity to drowsiness is an idiosyncratic factor which may also impact generalization. According to Van Dongen et al. (2003), the high variability in individual performance following sleep deprivation can be explained by the cognitive performance observed when the individual is not sleep-deprived. Van Dongen et al. (2003) also showed that individuals probably differ in their vulnerability to sleep deprivation, and that this is partially predictable from individual cognitive performance without deprivation, i.e. from the individual cognitive profile. Indeed, in driving simulator studies, drowsiness is often observed to develop in differing ways (Thiffault and Bergeron, 2003). Situational and personality factors, sleeping habits and driving history can contribute to the understanding of why some people fall asleep at the wheel while others do not. This points to the need to take into account drivers' traits or profiles when calibrating systems for the detection and prediction of driver fatigue.

5. Conclusion

In this study, different ANNs were used either to detect a drowsiness level or to predict when a driver's state will become impaired. The best models (those whose rates of successful detection or prediction are the highest) used information about eyelid closure, gaze and head movements and driving time. Performance on prediction is very promising, since the model can predict to within 5 min when the driver's state will become impaired. Moreover, modeling drowsiness as a continuum can lead to more precise detection systems offering refined results beyond simply detecting whether the driver is alert or drowsy. Future performance improvements could be achieved by using recurrent neural networks or dynamic neural networks to add temporality to the model, or adding other features like context information (traffic, type of road, weather etc.). These factors can influence the driver's state. However, as eyelid and head movements are difficult to record in a real car, the focus should be on improving a model using only driving performance, driving behavior (based on data provided by sensors in the car) and physiological measurements. Finally, a larger and more realistic dataset (far more subjects (wider range for age for example)), recorded in real, on-road, conditions (different times of the day for example) would be required to validate these models.

Conflict of interests

None.

Acknowledgments

This research was funded by a PhD grant from PSA Group via the OpenLab agreement with Aix-Marseille University and CNRS entitled "Automotive Motion Lab". We thank Marjorie Sweetko for correcting and improving the English manuscript, all the participants in this study, L. Marrou for helping to evaluate video recordings, P. Vars, V. Honnet, and M. Hing for their help with SCANeR® Software Developments.

References

Alhazmi, S., 2013. Towards Context-based Fatigue Detection System in Vehicular Area Network. University of Ottawa, Canada (Unpublished doctoral thesis).
Anund, A., Fors, C., Kecklund, G., Leeuwen, W.V., Åkerstedt, T., 2015. Countermeasures for Fatigue in Transportation: a Review of Existing Methods for Drivers on Road, Rail, Sea and in Aviation. Statens väg- Och Transportforskningsinstitut. (VTI report 852A).
Apparies, R.J., Riniolo, T.C., Porges, S.W., 1998. A psychophysiological investigation of the effects of driving longer-combination vehicles. Ergonomics 41 (5), 581–592.
Arnedt, J.T., Wilde, G.J., Munt, P.W., MacLean, A.W., 2001. How do prolonged wakefulness and alcohol compare in the decrements they produce on a simulated driving task? Accid. Anal. Prev. 33 (3), 337–344.
Beale, M., Hagan, M.T., Demuth, H.B., 1992. Neural Network Toolbox. Neural Network Toolbox, The Math Works 5. pp. 25.
Belz, S.M., Robinson, G.S., Casali, J.G., 2001. An on-Road investigation of commercial motor vehicle operator self assessment of fatigue as an indicator of driver fatigue. In: SAGE Publications Sage CA: Los Angeles, CA. Proceedings of the Human Factors and Ergonomics Society Annual Meeting Vol. 45. pp. 1576–1580.
Benoit, A., Caplier, A., 2005. Hypovigilence analysis: open or closed eye or mouth? Blinking or yawning frequency? In: IEEE Conference on Advanced Video and Signal Based Surveillance. AVSS 2005. pp. 207–212.
Bergasa, L.M., Nuevo, J., Sotelo, M.A., Barea, R., Lopez, M.E., 2006. Real-time system for monitoring driver vigilance. IEEE Trans. Intell. Transp. Syst. 7 (1), 63–77.
Besson, P., Bourdin, C., Bringoux, L., Dousset, E., Maiano, C., Marqueste, T., Vercher, J.-L., 2013. Effectiveness of physiological and psychological features to estimate helicopter pilots… workload: a bayesian network approach. IEEE Trans. Intell. Transp. Syst. 14 (4), 1872–1881.
Bhowmick, B., Chidanand Kumar, K.S., 2009. Detection and classification of eye state in IR camera for driver drowsiness identification. 2009 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) 340–345.
Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev. 44, 58–75.
Brown, I.D., 1997. Prospects for technological countermeasures against driver fatigue. Accid. Anal. Prev. 29 (4), 525–531.
Bundele, M.M., Banerjee, R., 2009. Detection of fatigue of vehicular driver using skin conductance and oximetry pulse: a neural network approach. In: Proceedings of the 11th International Conference on Information Integration and Web-based Applications & Services. New York, NY, USA : ACM. pp. 739–744.
Caffier, P.P., Erdmann, U., Ullsperger, P., 2003. Experimental evaluation of eye-blink parameters as a drowsiness measure. Eur. J. Appl. Physiol. 89 (3–4), 319–325.
Chauhan, A., Saroliya, A., Sharma, V., 2015. Design & Analysis of KNN algorithm for fatigue detection in vehicular drivers using Pulse Oximetry parameter. Int. J. Eng. Technol. Manage. 2 (3), 107–110.
Chen, J., Ji, Q., 2012. Drowsy driver posture, facial, and eye monitoring methods. In: Eskandarian, A. (Ed.), Handbook of Intelligent Vehicles. Springer, London, pp. 913–940.
Chen, R., 2013. Sitting Behaviour-based Pattern Recognition for Predicting Driver Fatigue. Deakin University, Australia (Unpublished doctoral thesis).
Daza, I.G., Bergasa, L.M., Bronte, S., Yebes, J.J., Almazán, J., Arroyo, R., 2014. Fusion of optimized indicators from Advanced Driver Assistance Systems (ADAS) for driver drowsiness detection. Sensors 14 (1), 1106–1131.
De Gennaro, L., Ferrara, M., Curcio, G., Cristiani, R., 2001. Antero-posterior EEG changes during the wakefulness?sleep transition. Clin. Neurophysiol. 112 (10), 1901–1911.
De Valck, E., De Groot, E., Cluydts, R., 2003. Effects of slow-release caffeine and a nap on driving simulator performance after partial sleep deprivation. Percept. Mot. Skills 96 (1), 67–78.
Dong, Y., Hu, Z., Uchimura, K., Murayama, N., 2011. Driver inattention monitoring system for intelligent vehicles: a review Intelligent Transportation Systems. IEEE Trans. 12 (2), 596–614.
Elsenbruch, S., Harnish, M.J., Orr, W.C., 1999. Heart rate variability during waking and sleep in healthy males and females. Sleep 22 (8), 1067–1071.
Eskandarian, A., Sayed, R., Delaigue, P., Blum, J., Mortazavi, A., 2007. Advanced Driver Fatigue Research. Federal Motor Carrier Safety Administration, Washington, DC Report: FMCSA-RRR-07–001.
Friedrichs, F., Yang, B., 2010. Drowsiness monitoring by steering and lane data based features under real driving conditions. In: Proceedings of the European Signal Processing Conference. Aalborg, Denmark. pp. 23–27.
Golding, J.F., 1998. Motion sickness susceptibility questionnaire revised and its relationship to other forms of sickness. Brain Res. Bull. 47 (5), 507–516.
Hajinoroozi, M., Mao, Z., Huang, Y., 2015. Prediction of driver's drowsy and alert states from EEG signals with deep learning. 2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) 493–496.
Hargutt, V., Kruger, H.-P., 2001. Eyelid movements and their predictive value for fatigue stages. In: Presented at the International Conference on Traffic and Transport Psychology ? ICTTP 2000. HELD 4–7 September 2000, Bern, Switzerland.
Healey, J.A., Picard, R.W., 2005. Detecting stress during real-world driving tasks using physiological sensors Intelligent Transportation Systems. IEEE Trans. 6 (2), 156–166.
Horne, J.A., Ostberg, O., 1975. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int. J. Chronobiol. 4 (2), 97–110.
Horne, J., Reyner, L., 1999. Vehicle accidents related to sleep: a review. Occup. Environ. Med. 56 (5), 289–294.
Ingre, M., Åkerstedt, T., Peters, B., Anund, A., Kecklund, Gör., 2006. Subjective sleepiness, simulated driving performance and blink duration: examining individual differences. J. Sleep Res. 15 (1), 47–53.
Ji, Q., Zhu, Z., Lan, P., 2004. Real-time nonintrusive monitoring and prediction of driver fatigue Vehicular Technology. IEEE Trans. 53 (4), 1052–1068.
Johns, M.W., 1991. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 14 (6), 540–545.
Ju, J.H., Park, Y.J., Park, J., Lee, B.G., Lee, J., Lee, J.Y., 2015. Real-Time driver's biological signal monitoring system. Sens. Mater. 27 (1), 51–59.
Kaida, K., Åkerstedt, T., Kecklund, G., Nilsson, J.P., Axelsson, J., 2007. Use of subjective and physiological indicators of sleepiness to predict performance during a vigilance task. Ind. Health 45 (4), 520–526.
Karrer, K., Vöhringer-Kuhnt, T., Baumgarten, T., Briest, S., 2004. The role of individual differences in driver fatigue prediction. In: Third International Conference on Traffic and Transport Psychology. Nottingham, UK. pp. 5–9.
Krajewski, J., Batliner, A., Golz, M., 2009a. Acoustic sleepiness detection: framework and validation of a speech-adapted pattern recognition approach. Behav. Res. Methods 41 (3), 795–804.
Krajewski, J., Sommer, D., Trutschel, U., Edwards, D., Golz, M., 2009b. Steering wheel behavior based estimation of fatigue. Proceedings of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design 118–124.
Lal, S.K., Craig, A., 2001. A critical review of the psychophysiology of driver fatigue. Biol. Psychol. 55 (3), 173–194.
Larue, G.S., 2010. Predicting effects of monotony on driver's vigilance. Centre for Accident Research and Road Safety. Queensland University of Technology, Australia (Unpublished doctoral thesis).
Lee, B.G., Chung, W.-Y., 2012. Driver alertness monitoring using fusion of facial features and bio-Signals. IEEE Sens. J. 12 (7), 2416–2422.
Lee, J.D., Fiorentino, D., Reyes, M.L., Brown, T., Ahmad, O., Fell, J., Dufour, R., 2010. Assessing the Feasibility of Vehicle-based Sensors to Detect Alcohol Impairment 811. National Highway Traffic Safety Administration, Washington, DC,DOT HS, pp. 358.
Lee, B.L., Lee, B.G., Chung, W.Y., 2016. Standalone wearable driver drowsiness detection system in a smartwatch. IEEE Sens. J. 16 (13), 5444–5451.
Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares. Q. Appl. Math. 2 (2), 164–168.
Li, L., Werber, K., Calvillo, C.F., Dinh, K.D., Guarde, A., König, A., 2014. Multi-Sensor soft-Computing system for driver drowsiness detection. In: Snášel, V., Krömer, P., Köppen, M., Schaefer, G. (Eds.), Soft Computing in Industrial Applications. Springer International Publishing, pp. 129–140.
Liang, Y., Reyes, M.L., Lee, J.D., 2007. Real-Time detection of driver cognitive distraction using support vector machines. IEEE Trans. Intell. Transp. Syst. 8 (2), 340–350.
Liu, C.C., Hosking, S.G., Lenné, M.G., 2009. Predicting driver drowsiness using vehicle measures: recent insights and future challenges. J. Saf. Res. 40 (4), 239–245.
Marin-Lamellet, C., Paire-Ficout, L., Lafont, S., Amieva, H., Laurent, B., Thomas-Antérion, C., Fabrigoule, C., 2003. Mise En Place d'un Outil d'évaluation Des déficits Attentionnels Affectant Les Capacités De Conduite Au Cours Du Vieillissement Normal Et Pathologique: L'étude SÉROVIE 81. Recherche – Transports – Sécurité, pp. 177–189.
McDonald, A.D., Lee, J.D., Schwarz, C., Brown, T.L., 2013. Steering in a random forest ensemble learning for detecting drowsiness-Related lane departures. Hum. Factors J. Hum. Factors Ergon. Soc (18720813515272).
Murata, A., Naitoh, K., 2015. Multinomial logistic regression model for predicting driver's drowsiness using only behavioral measures. J. Traffic Trans. Eng. 3, 80–90.
Murata, A., Ohta, Y., Moriwaka, M., 2016. Multinomial logistic regression model by stepwise method for predicting subjective drowsiness using performance and behavioral measures. In: In: Goonetilleke, R., Karwowski, W. (Eds.), Advances in Physical Ergonomics and Human Factors 489. Springer International Publishing, Cham, pp.

665–674.
Peiris, M.T.R., Jones, R.D., Davidson, P.R., Carroll, G.J., Signal, T.L., Parkin, P.J., Bones, P.J., 2005. Identification of vigilance lapses using EEG/EOG by expert human raters. 2005 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 1–7, 5735–5737.
Philip, P., Taillard, J., Guilleminault, C., Quera, S., Bioulac, B., Ohayon, M., 1999a. Long distance driving and self?induced sleep deprivation among automobile drivers. Sleep 22 (4), 475–480.
Philip, P., Taillard, J., Quera-Salva, M., Bioulac, B., Åkerstedt, T., 1999b. Simple reaction time, duration of driving and sleep deprivation in young versus old automobile drivers. J. Sleep Res. 8 (1), 9–14.
Philip, P., Taillard, J., Sagaspe, P., Valtat, C., Sanchez-Ortuno, M., Moore, N., Bioulac, B., 2004. Age, performance and sleep deprivation. J. Sleep Res. 13 (2), 105–110.
Philip, P., Sagaspe, P., Taillard, J., Valtat, C., Moore, N., Åkerstedt, T., Bioulac, B., 2005. Fatigue, sleepiness, and performance in simulated versus real driving conditions. Sleep 28 (12), 1511.
Rebolledo-Mendez, G., Reyes, A., Paszkowicz, S., Domingo, M.C., Skrypchuk, L., 2014. Developing a body sensor network to detect emotions during driving. IEEE Trans. Intell. Transp. Syst. 15 (4), 1850.
Reimer, B., Coughlin, J.F., Mehler, B., 2009. Development of a driver aware vehicle for monitoring, managing & motivating older operator behavior. Proceedings of the ITS-America 1–9.
Riemersma, J.B.J., Sanders, A.F., Wildervanck, C., Gaillard, A.W., 1977. Performance decrement during prolonged night driving. Vigilance. Springer, pp. 41–58.
Rodriguez Ibañez, N., García González Á, M., Ramos Castro, J.J., Fernández Chimeno, M., 2011. Drowsiness detection by thoracic effort signal snalysis with professional drivers in real environments. Driver Distraction & Inattention 2011: Program, Presentations & Reviewed Papers.
Rossi, R., Gastaldi, M., Gecchele, G., 2011. Analysis of driver task-related fatigue using driving simulator experiments. Proc. Soc. Behav. Sci. 20, 666–675.
Rost, M., Zilberg, E., Xu, Z.M., Feng, Y., Burton, D., Lal, S., 2015. Comparing contribution of algorithm based physiological indicators for characterisation of driver drowsiness. J. Med. Bioeng. 4 (5), 391–398.
Samiee, S., Azadi, S., Kazemi, R., Nahvi, A., Eichberger, A., 2014. Data fusion to develop a driver drowsiness detection system with robustness to signal loss. Sensors 14 (9), 17832 (14248220).
Sayed, R., Eskandarian, A., 2001. Unobtrusive drowsiness detection by neural network learning of driver steering. Proceedings of The Institution of Mechanical Engineers Part D-Journal of Automobile Engineering 215 (9), 969–975.
Shahid, A., Wilkinson, K., Marcu, S., Shapiro, C.M., 2011. Karolinska sleepiness scale (KSS). In: Shahid, A., Wilkinson, K., Marcu, S., Shapiro, C.M. (Eds.), STOP, THAT and One Hundred Other Sleep Scales. Springer, New York, pp. 209–210 (ch 47).
Stein, P.K., Pu, Y., 2012. Heart rate variability, sleep and sleep disorders. Sleep Med. Rev. 16 (1), 47–66.
Sukanesh, R., Vijayprasath, S., 2013. Certain investigations on drowsiness alert system based on heart rate variability using LabVIEW. WSEAS Trans. Inf. Sci. Appl. 10 (11).
Tango, F., Calefato, C., Minin, L., Canovi, L., 2009. Moving attention from the road: a new methodology for the driver distraction evaluation using machine learning approaches. 2nd Conference on Human System Interactions 2009, 596–599 (HSI '09).
Thiffault, P., Bergeron, J., 2003. Fatigue and individual differences in monotonous simulated driving. Personality Individual Differences 34 (1), 159–176.
Torkkola, K., Gardner, M., Schreiner, C., Zhang, K., Leivian, B., Zhang, H., Summers, J., 2008. Understanding driving activity using ensemble methods. In: Prokhorov, D. (Ed.), Computational Intelligence in Automotive Applications. Springer, Berlin Heidelberg, pp. 39–58.
Van Dongen, H.P.A., Rogers, N.L., Dinges, D.F., 2003. Sleep debt: theoretical and empirical issues. Sleep Biol. Rhythms 1 (1), 5–13.
Van Dongen, H.P.A., Baynard, M.D., Maislin, G., Dinges, D.F., 2004a. Systematic inter-individual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep 27 (3), 423–433.
Van Dongen, H.P.A., Maislin, G., Dinges, D.F., 2004b. Dealing with inter-individual differences in the temporal dynamics of fatigue and performance: importance and techniques. Aviat. Space Environ. Med. 75 (3), A147–A154.
Verwey, W.B., Zaidel, D.M., 2000. Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behaviour. Personality Individual Differences 28 (1), 123–142.
Wang, X., Xu, C., 2016. Driver drowsiness detection based on non-intrusive metrics considering individual specifics. Accid. Anal. Prev. 95, 350–357 (Part B).
Watson, A., Zhou, G., 2016. Microsleep prediction using an EKG capable heart rate monitor. 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE) 328–329.
Wesensten, N.J., Belenky, G., Thorne, D.R., Kautz, M.A., Balkin, T.J., 2004. Modafinil vs. caffeine: effects on fatigue during sleep deprivation. Aviat. Space Environ. Med. 75 (6), 520–525.
Wierwille, W.W., Ellsworth, L.A., 1994. Evaluation of driver drowsiness by trained raters. Accid. Anal. Prev. 26 (5), 571–581.
Yang, G., Lin, Y., Bhattacharya, P., 2010. A driver fatigue recognition model based on information fusion and dynamic Bayesian network. Inf. Sci. 180 (10), 1942–1954.
Yeo, M.V.M., Li, X., Shen, K., Wilder-Smith, E.P.V., 2009. Can SVM be used for automatic EEG detection of drowsiness during car driving? Saf. Sci. 47 (1), 115–124.
Zhang, Y., Owechko, Y., Zhang, J., 2004. Driver cognitive workload estimation: a data-driven perspective. The 7th International IEEE Conference on Intelligent Transportation Systems, 2004. Proceedings 642–647.