Predicting Signs of Drowsiness from Drivers Blinking and Driving Behavior

Jinseok Moon[1], Junya Morita[1,2], Yu Enokibori[1], Takatsugu Hirayama[1,3], Albert Ali Salah[1,4], Kenji Mase[1], Yoshiyuki Hatakeyama[5] and Hirotaka Kaji[5]

[1]: Graduate School of Informatics, Nagoya University, Japan
[2]: Department of Behavior Informatics, Shizuoka University, Japan
[3]: Institutes of Innovation for Future Society, Nagoya University, Japan
[4]: Computer Engineering Dept, Bogazici University, Turkey
[5]: Z-Frontier Div. TOYOTA MOTOR CORPORATION
E-mail: mun@cmc.is.i.nagoya-u.ac.jp

Abstract The necessity of detecting and preventing drowsiness during driving is increasing, because drowsy driving is one of the important factors of fatal crashes. The American Automobile Association Foundation for Traffic Safety reported that drowsy driving accounted for 6% of total crashes and 21% of total fatal crashes in the USA. In this work, we focus on detecting signs of drowsiness. Knowing signs of drowsiness makes it possible to inform the driver of drowsiness in advance. We propose an algorithm to detect drowsiness signs from drivers blinking and driving behavior. Using these, we search for a parameter set suitable for drowsiness prediction using Support Vector Machines. We also train a drowsiness sign detection model for drowsiness intensity and compare its performance for weak and strong drowsiness intensity groups. In the prediction experiments, we obtained average F-scores in the range of 0.441 to 0.748 for weak drowsiness, and 0.432 to 0.757 for strong drowsiness, respectively. Our results show that driving behavior is a good predictor of drowsiness, and it is possible to successfully detect drowsiness while driving. However, our results also show that adding the blink data to the prediction model as a feature degrades the performance compared to the original model that uses only driving behavior.

Keywords driver drowsiness, CAN-data, eye blink, drowsiness prediction, driving behaviour, machine learning, fatigue

1. INTRODUCTION

Drowsy driving is one of the important factors of fatal crashes. According to the National Highway Traffic Safety Administration [1], there were about 5.9 million traffic accidents in 2005-2009 in the USA, and drowsy driving was involved in 83 thousand of these cases, representing 1.4% of accidents. The report was collected by the police, but there are also accidents caused by drowsy driving which are not officially accounted for. The American Automobile Association Foundation for Traffic Safety [2] reported that drowsy driving accounted for 6% of total crashes and 21% of total fatal crashes in the USA. As described above, traffic accidents caused by drowsy driving occur frequently, and result in fatalities in many cases, so the necessity of preventing drowsy driving is increasing.

Detecting signs of drowsiness before the driver gets sleepy can help us alert the drivers to the risk of continuing driving, and to guide them to a safe place. Biological signals such as driver's head movements, facial expressions, brain waves, blinks, etc. can serve as indicators of drowsiness. In this research, we propose an algorithm to detect drowsiness signs using driving behavior related to steering and accelerator, as well as blinking patterns of the driver. Our aim is to examine the possibility of timely and actionable detection of drowsiness signs, and to understand the contribution of each feature to this goal.

2. RELATED RESEARCH

There are many works to detect drowsiness while driving, advocating various methods. Other concepts like "fatigue," or "arousal level" are used interchangeably with "drowsiness" in these approaches. The features used can be divided into three groups, namely, biological signals, driver's bodily behaviors, and driving behavior, respectively:

- Biological signals: Eye blinks [3, 4, 5, 6], brain waves [7, 8].
- Driver's bodily behaviors: changes in the movement of the face or the head [4, 9].
- Driving behavior: standard deviation of lane position (SDLP), steering, and vehicle speed [3, 10].

There is some research to detect the driver's drowsiness on time. He et al. [3] detected blinks using the Google Glass infrared proximity sensor, from sudden changes in the value of the sensor at the moment of blinking. Also, during the driving simulation, drowsiness was detected by using decreased lane maintaining ability, the increased frequency of blinking, and the extension of the braking response time. Yeo et al. [7] proposed the auto drowsiness detector from brain waves using

Support Vector Machine (SVM). They used the usual driving video for driving simulation. Using SVM with 2 label; alert or drowsy state, they succeed to classify at 99.3%. Ji et al. [4] developed a probabilistic model of human fatigue based on visual cues. They measured drowsiness in real-time, from a probabilistic model using Bayesian Network, with facial recognition, head movement, eye blinking, PERCLOS (PERcentage of eye CLOSure), and contextual information. They also showed a correlation between PERCLOS and drowsiness, and found that the PERCLOS of drowsy state is higher compared to the awake state. Sayed and Eskandarian [10] used Artificial Neural Networks (ANNs) with only the steering data as a feature for drowsiness detection. They report a high performance, and the steering data are easy to obtain, without recourse to image processing or measuring biological signals.

Also, there is some research to predict drowsiness in advance. Vural et al. [9] detected driver's drowsiness by machine learning using steering in addition to facial muscle and head movement analysis. As a result of successful application of strong classifiers such as Adaboost or multinomial ridge regression, the classification accuracy was about 90%. In addition, they showed that information collected one minute before an accident occurred could be used reliably as a predictor of drowsiness. Sato et al. [6] indexed the driver's arousal level while automatic driving, using PERCLOS and Vestibulo-Ocular Reflex (VOR). Their results showed that in the event which VOR could be acquired, awareness of drowsiness can be detected in advance. Hatakeyama [5] showed the effectiveness of information related to open eyes and closed eyes in the prediction of drowsiness. Hatakeyama proposed the hypothesis that when people become drowsy, there is a positive or negative change in the standard deviation of the eye opening time (SDEOP). In addition, Hatakeyama adopted the 5-level model of drowsiness proposed by Kitajima et al. [11]. The drowsiness prediction logic was built by using these drowsiness levels, and relied on the changes in the open-eye time. However, the author noted that drowsiness prediction accuracy is lower for out-of-training conditions.

While there is significant literature for both detecting and predicting driver's drowsiness, there are few studies that investigated how much in advance of a truly dangerous state it is possible to predict drowsiness. Also, the relative importance of different features in this problem is not established. Our study focuses on these points and investigates how far ahead we can predict drowsiness and which features are more important for drowsiness prediction.

3. PROPOSED METHOD

3.1. Dataset

We took five round trips data on the real highway about one hour each way, for each of five subjects (20-40 years old, four men and one woman), 50 trip samples in total. The route was the section of about 100km between Fuji and Hamamatsu, Tomei and Shintomei highway. While each sample is acquired from the same road, the length of data samples fluctuates, based on road conditions, etc. There are also some instances where the experiment was stopped without finishing the route, due to the condition of the subjects. And also, for the safety of driver, there was a supervisor behind the driver during the experiment. Appropriate subjects with a certain driving ability were recruited for the experiment. The design of the experiment was approved and conducted according to the "Ethical Guidelines for Research Involving with Human Subjects" of Toyota Motor Corporation. Informed consent was obtained from all participants, and they were informed about all parts of the experiment including details of the experimental procedures and our privacy protection policies.

Our dataset includes Controller Area Network (CAN) data of vehicles and time series data of driver's eye blinking. CAN data was taken at 100Hz, and includes the speed (km/h), the status of the brake lamp, the amount of press for the brake and accelerator pedals, the steering wheel turning degree and the degree of the zero position. Eye blinking data was taken at 30Hz. We acquired the driver's face video while driving. After the experiment, we annotated all the instances when the driver closed his or her eyes. Also, for 50 samples of data, three operators judged whether the driver became drowsy or not, both by watching facial videos, and by the time when becoming drowsy was noticed during driving, followed by the stopping of the car. The cases when the car should be stopped because of drowsiness are labeled as STOP-data, whereas the cases where the experiment could be continued are labeled as NONSTOP-data. Because annotation was done after driving, there are some STOP-data which experiment continued. But, when driver reported self-awareness of drowsiness, the experiment was stopped. Of the 50 samples we collected, 30 are labeled as STOP-data and 20 are NONSTOP-data. We further labeled the drowsiness in the STOP-data as "weak" or "strong", depending on the duration of drowsiness.

3.2. Data Preprocessing

3.2.1. CAN data

Among the CAN data, we used the accelerator pedal amount and the steering wheel turning angle. The speed data included in the CAN was excluded, because it has a high correlation with stepping on the accelerator. When considering the environment of the expressway, braking events were rare, and consequently, we excluded the brake data as well. Krajewski et al. [12] and Wang et al. [13] showed that when the driver gets drowsy, steering wheel changes are coarser, and acceleration changes are more significant compared to the alert condition. We took the variance V_Y from the difference data Y of 10 seconds as a feature (X_i is the sampled data, and n is the window size). When the driver gets drowsy, more fluctuation may appear and variance is supposed to increase.

$$Y_i = X_{i+1} - X_i, (i = 1, \ldots, n) \tag{1}$$

$$V_Y = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{2}$$

3.2.2. *Blink data*

In many works, blink rate and PERCLOS (PERcentage of Eyelid CLOSure) are widely used for driver's drowsiness detection. PERCLOS is defined as the percentage of eye closing in a specific time period (e.g. one minute). Blink rate is also defined within a specific time period (e.g. number of times per minute). These two features are known to have a high correlation with the driver's drowsiness.

3.2.3. *Feature definitions*

In order to predict the drowsiness, it is necessary to decide a section considered to be related to drowsiness from data. We aim to detect drowsiness a certain interval of time before it happens, and we call this interval the *Gap*. The area where we actually watch (call *Feature*) was extracted from a time window *F* are used to predict drowsiness at time *t* + *G*, where *G* is the duration of the *Gap*, and *t* is time of the end of *F*. Adding the intensity *I* of the drowsiness, we parameterize our Drowsiness models with the triple *D (I, F, G)*. For drowsiness intensity *I*, two levels, namely, weak and strong, are used as shown in Fig. 1.

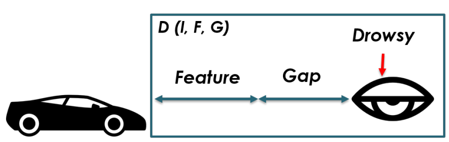


Fig. 1: The definition of *D (I, F, G)*. *I* denotes the drowsiness intensity, *F* the length of the feature extraction period, *G* the gap before the occurrence of drowsiness.

3.2.4. *Feature Extraction*

When the data are extracted from windows of 10 seconds, the dimension of the input vector is too large compared to the number of data items in our training set. In order to prevent over-learning, we have reduced the dimension of the features. We split our input into six windows, and the maximum width of one window is set to 60 seconds like Table. 1. When the window width exceeds 60 seconds, we add more windows. The selected features $\mathbf{X}$ are represented as given in Eq. 3 and Eq. 4.

$$\mathbf{X}_i = (acceleration, steering, perclos, blinkrate) \tag{3}$$

$$\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n) \tag{4}$$

Table 1: Feature extraction and preprocessing. This is the example of *F*=120s. After split the input, get average in the same window.

Features / Time	10	20	30	40	50	60	70	80	90	100	110	120
Acceleration	Avg		Avg		Avg		Avg		Avg		Avg	
Steering	Avg		Avg		Avg		Avg		Avg		Avg	
PERCLOS	Avg		Avg		Avg		Avg		Avg		Avg	
Blink Rate	Avg		Avg		Avg		Avg		Avg		Avg	

3.3. Proposed Method

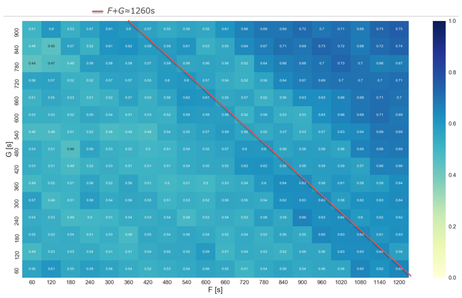
We trained a drowsiness sign detection model for drowsiness intensity and compared its performance for weak and strong drowsiness intensity groups. Using the driver's blinking and driving behavior, we searched for a parameter set suitable for drowsiness prediction using Support Vector Machines, while varying the *F* and *G* of *D (I, F, G)*. *F* was varied between 1 minute to 20 minutes, and *G* was varied between 1 minute to 15 minutes. Both *F, G* are varied in increments of one minute. In our dataset, we use positive drowsiness samples from the first *F* section from *D (I, F, G)* of the STOP-data, and the negative samples are taken randomly from the NONSTOP-data with the same length of *F* of *D (I, F, G)*. The number of positive and negative samples were selected to be equal, but we caution the reader that this is ordinarily an unbalanced problem, with the prior probability of the presence of drowsiness much lower than its absence. We used strong drowsiness data for training the strong drowsiness model. For weak drowsiness model, we used both weak and strong data. To detect a weak drowsiness, strong drowsiness data are considered also.

We evaluated the drowsiness model by using 10 folds cross-validation. Also, we automatically set the best parameters for the model with grid search. Finally, we calculated the F-score of each model of *D (I, F, G)*. The F-score was calculated by using Eq. 5.

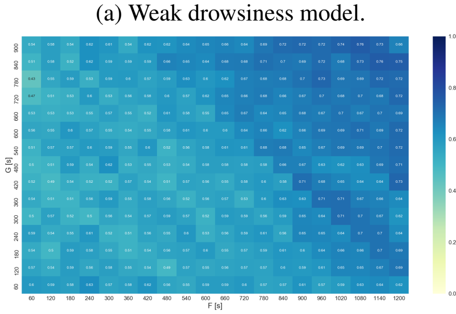
$$\textit{F-score} = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{5}$$

4. RESULTS

We calculated the F-score of each weak and strong drowsiness model, by parameterizing the length of *F* and *G*. As can be seen in Table. 2, for weak drowsiness, the F-scores range from 0.441 to 0.748 ($M = 0.577, SD = 0.064$), whereas for strong drowsiness, they range from 0.432 to 0.757 ($M = 0.608, SD = 0.064$). The entire F-score range for each model is presented in Fig. 2. Because negative samples were taken randomly, we evaluated the model 10 times with different sets, and calculated the average F-score of each model with parameterized *D (I, F, G)*.



(a) Weak drowsiness model.



(b) Strong drowsiness model.

Fig. 2: F-score of drowsiness sign detection model. The x-axis denotes the length of *F*, and the y-axis denotes the length of *G*. The F-scores are in the range of *0-1* and are presented with yellow-green-blue colors. In the weak model, the red line presents the area where *F+G* is about 1260s.

Table 2: As evaluation of the models, we calculated F-scores from the results of classification. I_{weak} and I_{strong} mean weak drowsiness model and strong drowsiness model.

Model / F-score	Min	Max	Average	Standard Deviation
I_{weak}	0.441	0.748	0.577	0.064
I_{strong}	0.432	0.757	0.608	0.064

In the weak drowsiness model, when the *Total Time* of *F+G* is 1200s, and 1800s, there were the specific slanted strap-like results which F-score was high. We examined correlations using linear regression, and calculated the correlation coefficients. Both models had a similar trend where when the *Total Time* became higher, the F-scores were also higher, as shown in Table. 3 and Fig. 3. We discuss these results more in Section. 5.

We have also compared the performance between weak and strong models in Fig. 4. As can be seen, the strong model's score was a little higher than the weak model.

In Fig. 5, we investigate which feature has a stronger effect. The score was the highest when we used only features from driving behavior (steering and acceleration). When combined with blinking, the results were similar to using only driver's blinking.

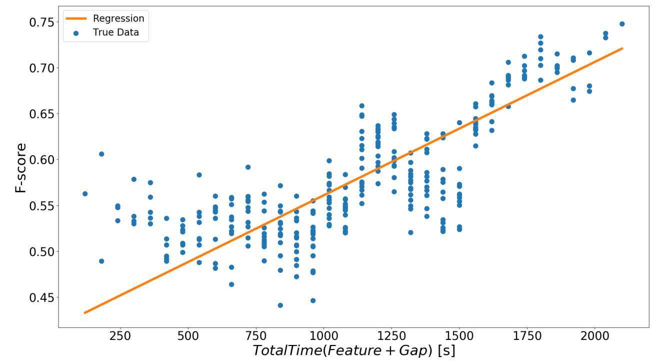


Fig. 3: The regression between F-score and *Total Time* of the weak drowsiness model. There is an increasing trend between F-score and *Total Time*.

Table 3: Result of linear regression between F-score and *Total Time* and between *F* and *G*. *** means $p < .001$. The correlation coefficients are predicted from regression.

Intensity	explanatory variable	R^2	*F-statistic*
weak	*Total Time*	0.590	430.7***
	G, F	0.620	244.4***
strong	*Total Time*	0.658	573.0***
	G, F	0.709	364.7***

5. DISCUSSION

We have proposed a generic drowsiness prediction system that uses driver's driving behavior and blinking. In the prediction experiments, we have obtained F-score maximum 0.748 for weak drowsiness, and 0.757 for strong drowsiness, respectively.

Liu et al. [14] state that when the predicting time window is too short, it can be treated as noise, but when it is too long, it is not appropriate, because it's too slow to predict drowsiness. In our work as well, as the *F* and the *G* features are extracted from longer intervals we can observe an increasing trend in accuracy. The feature dimensionality is naturally increased, and this needs to be counterbalanced with either more training samples, or via dimensionality reduction. However, through the basic tendency of accuracy increase, it is possible to discern some dips. We postulate that the tendency of increase is due to physiological features (e.g. longer driving, become more drowsy), but there is also some effect of the *F* parameterization. Additionally, it is necessary to investigate what the 'peak-peak interval' corresponds to, and to consider the method of exclusion (or compensation) of the influence of physiological characteristics. Through the linear regression

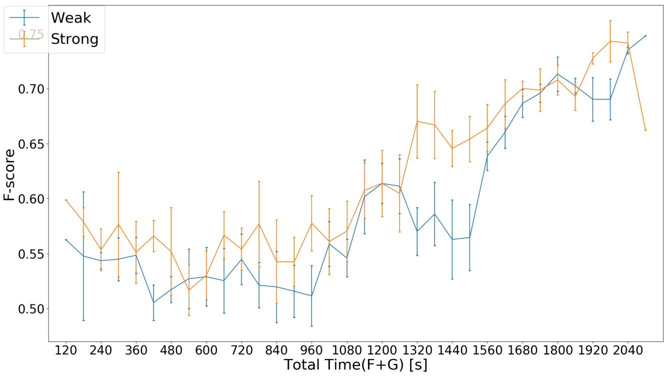


Fig. 4: The averaged F-score of each model. The x-axis denotes the *Total Time* and the y-axis denotes the F-score. The strong drowsiness model's performance is better than the weak model in some instances (e.g. 420s, 1360s).

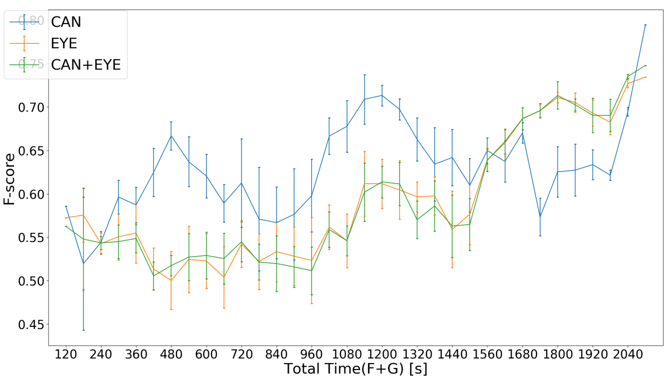


Fig. 5: The comparison of different feature sets. Using only the CAN data, there are some peaks at 480s, 1200s and 1680s. When we combine the driver's blinking and driving behaviors, the trend becomes similar to the system which used only driver's blinking.

of F-score and *Total Time* as shown in Fig. 3, the regression line illustrates the correlation between time of *F+G* and the prediction performance. In the future, contextual information should be integrated into the model to remove the effect of the physiological factors.

The results shown in Fig. 5 indicate that adding the blink data to the prediction model as a feature degrades the performance compared to the original model that uses only driving behavior. We have used generic models for blinking patterns, but the literature states that the blinking rate can either go up, or down, as stated before. It may be useful to train subject-specific models to investigate how blinking patterns can contribute to accuracy. At the moment, we do not have sufficient data to explore this. Sayed and Eskandarian [10] changed the raw steering data into a vector, and got high performance of drowsiness detection. Fairclough [15] also used the mean steering wheel reversal rate, instead of the raw vehicle data. In the present system, we take the difference of sequences and use the variance of the difference for CAN data preprocessing. There is room for improvement in feature extraction from the CAN data, and semantically meaningful features, such as steering wheel reversals or mean amplitude of steering wheel movements could provide additional information and improvements in discrimination performance.

6. CONCLUSION

We have examined whether it is possible to detect drowsiness signs from driver's blinking patterns and driving behavior. We distinguished between weak drowsiness and strong drowsiness, where the latter is more dangerous, but easier to detect as a category. Our proposed approach could successfully predict drowsiness signs for unlearned data and eventually predict drowsiness.

The driving behaviors, like steering or acceleration, can be combined with contextual information for more accurate predictions (e.g. the traffic jam or the course of experiment). However, it is not clear whether the additional accuracy will justify the increasing computational cost and system cost. Based on our results, we propose that it is possible to predict drivers drowsiness using only the drivers driving behavior and blinking (i.e. without contextual information).

In our study, longer feature extraction windows resulted in better accuracy, but how much of this increase results due to physiological characteristics is not assessed. In the future, these factors could be further explored and dissociated to get a better understanding of drowsiness progression.

7. REFERENCES

[1] National Center for Statistics and Analysis, "Traffic safety facts: Drowsy driving (crashstats brief statistical summary. report no. dot hs 811 449)," , Washington, DC: National Highway Traffic Safety Administration, Mar 2011.

[2] Brian C. Tefft, "Prevalence of motor vehicle crashes involving drowsy drivers, united states, 2009-2013," , AAA Foundation for Traffic Safety, Nov 2014.

[3] J He, W Choi, Y Yang, J Lu, X Wu, and K Peng, "Detection of driver drowsiness using wearable devices: A feasibility study of the proximity sensor," *Applied ergonomics*, vol. 65, pp. 473–480, 2017.

[4] Q Ji, Z Zhu, and P Lan, "Real-time nonintrusive monitoring and prediction of driver fatigue," *IEEE Transactions on vehicular technology*, vol. 53, no. 4, pp. 1052–1068, 2004.

[5] Y Hatakeyema, "Feasibility study of drowsy driving prediction based on eye opening time," Tech. Rep., SAE Technical Paper, 2017-01-1398, 2017.

[6] K Sato, S Yoshida, H Ishida, Y Hirata, J Emoto, M Omae, G Abe, and N Uchida, "Changes in driver's levels of wakefulness during automatic driving in an actual driving environment," in *JSAE Annual Spring Congress, PACIFICO YOKOHAMA*. May 23rd, 2018.

[7] Mervyn VM Yeo, Xiaoping Li, Kaiquan Shen, and Einar PV Wilder-Smith, "Can svm be used for automatic eeg detection of drowsiness during car driving?," *Safety Science*, vol. 47, no. 1, pp. 115–124, 2009.

[8] SKL Lal, A Craig, P Boord, L Kirkup, and H Nguyen, "Development of an algorithm for an eeg-based driver fatigue countermeasure," *Journal of safety Research*, vol. 34, no. 3, pp. 321–328, 2003.

[9] E Vural, M Çetin, A Erçil, G Littlewort, M Bartlett, and J Movellan, "Machine learning systems for detecting driver drowsiness," in *In-vehicle Corpus and Signal Processing for Driver Behavior*, pp. 97–110. Springer, 2009.

[10] R Sayed and A Eskandarian, "Unobtrusive drowsiness detection by neural network learning of driver steering," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 215, no. 9, pp. 969–975, 2001.

[11] H Kitajima, N Numata, K Yamato, and Y Goi, "Prediction of automobile driver sleepiness," *The Japan Society of Mechanical Engineers Collected Papers (C book)*, vol. 63, no. 613, pp. 93–100, 1997.

[12] J Krajewski, D Sommer, U Trutschel, D Edwards, and M Golz, "Steering wheel behavior based estimation of fatigue," in *Driving Assessment 2009: 5th International Driving Symposium on Human Factors in Driving Assessment, Training and Vehicle Design*. Big Sky MT, United States, 2009.

[13] Qiong Wang, Jingyu Yang, Mingwu Ren, and Yujie Zheng, "Driver fatigue detection: a survey," in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*. IEEE, 2006, vol. 2, pp. 8587–8591.

[14] CC Liu, SG Hosking, and MG Lenné, "Predicting driver drowsiness using vehicle measures: Recent insights and future challenges," *Journal of safety research*, vol. 40, no. 4, pp. 239–245, 2009.

[15] SH Fairclough and R Graham, "Impairment of driving performance caused by sleep deprivation or alcohol: a comparative study," *Human factors*, vol. 41, no. 1, pp. 118–128, 1999.