Real-driving-implementable drowsy driving detection method using heart rate variability based on long short-term memory and autoencoder

Hiroki Iwamoto * **Kentaro Hori** ** **Koichi Fujiwara** *** **Manabu Kano** ****

* *Department of Systems Science, Kyoto University, Kyoto, Japan (e-mail: iwamoto.hiroki.38m@st.kyoto-u.ac.jp).*
** *Quadlytics inc., Kyoto, Japan (e-mail: hori.kentaro@quadlytics.com).*
*** *Department of Material Process Engineering, Nagoya University, Nagoya, Japan (e-mail: fujiwara.koichi@hps.material.nagoya-u.ac.jp).*
**** *Department of Systems Science, Kyoto University, Kyoto, Japan (e-mail: manabu@human.sys.i.kyoto-u.ac.jp).*

Abstract: Drowsy driving is a fatal problem that may cause traffic accidents. Although many driver drowsiness detection methods have been proposed, most of them have problems in input data availability, robustness to real driving environments, or detection precision. A drowsiness detection method based on heart rate variability (HRV), which is an R-R interval (RRI) fluctuation obtained from an electrocardiogram (ECG), has been proposed since ECG is easy to measure by using a wearable sensor. HRV is related to the autonomic nervous system (ANS) and is affected by drowsiness. However, its drowsiness detection performance was not always satisfactory. This study proposes a new drowsiness detection method using raw RRI data instead of HRV to improve the drowsiness detection performance. The proposed method uses raw RRI time series as inputs, and a drowsiness detection model is trained based on long short-term memory (LSTM) and autoencoder (AE), which are types of neural networks. RRI data during driving were collected from 25 participants using a driving simulator. The drowsiness detection model was trained following the proposed method. The experimental result showed that the proposed method achieved an AUC of 0.88, a sensitivity of 81%, and a specificity of 91%, which was higher than the HRV-based method. The result suggests that it is better to use raw RRIs as inputs than HRV features.

Keywords: Driver drowsiness, Anomaly detection, Heart rate variability, Long short-term memory, Autoencoder

1. INTRODUCTION

Drowsiness in driving may cause fatal traffic accidents; thus many studies have tried to detect driver drowsiness. Most of these studies classified into three groups following input data types: video, electroencephalogram (EEG), and vehicle behavior.

Video-based methods mainly utilize blinks (Rahman et al., 2015), facial expressions (Chiou et al., 2020), or a combination of them (Zhang et al., 2019). They are commonly used because drivers do not have to wear any intrusive measuring devices. However, these methods sometimes require strict light conditions for adequate video recording, and it is difficult to detect drowsiness in a consistent performance in real driving environments. EEG-based methods (e.g. Lin et al. (2010)) are also common since EEG measurement is the gold standard for sleep scoring in sleep medicine. Although many EEG-based studies reported high detection performances in experimental environments, EEG can rarely be used in real driving since EEG is intolerant of motion artifacts generated by vibration from engines and roads and electromagnetic environments in vehicles. Behavior-based methods mainly utilize driving data such as steering activity (Arefnezhad et al., 2019) and lateral position variability (Ingre et al., 2006). Although these methods are non-intrusive, they function in only specific situations because of some limitations, such as course shapes. Many drowsiness detection methods proposed ever have problems for implementation in real driving, and other types of data are required to realize a practical driver drowsiness detection system.

Heart rate variability (HRV), which is an R-R interval (RRI) fluctuation in an electrocardiogram (ECG), has the potential to be such a real-driving-implementable type of data. HRV is known to be affected by the autonomic nervous system (ANS) and may be useful for sleep monitoring (Malik et al., 1996). In addition, since the amplitude of R waves in ECG signals is much higher than EEG signals, it is much easier to measure RRI data than EEG. Thus, HRV-based drowsiness detection may be more feasible in real driving. In addition, various wearable RRI devices such as intelligent T-shirts (cf. Wu et al. (2019)) have been developed, and HRV analysis has been widely used in the health monitoring field.

In HRV analysis, HRV features are commonly used. In the context of driver drowsiness detection, Fujiwara et al. (2019) utilized eight features, such as RRI mean, RRI standard deviation, and RRI power of certain frequency bands. In their method, awake HRV data and drowsy HRV data are respectively regarded as normal and abnormal, and driver drowsiness was detected by multivariate statistical process control (MSPC), an anomaly detection method based on principal component analysis (PCA). However, they manually select input HRV features in a trial-and-error manner.

Deep learning (DL) may reduce the burden of HRV feature selection and improve drowsiness detection performance since DL can automatically extract appropriate features from input data when a model is trained. In addition, HRV feature extraction may lose some important information for drowsiness detection. Thus, using raw RRIs as DL input instead of HRV features may contribute to realizing a practical driver drowsiness detection system. Iwasaki et al. (2021) showed that the sleep apnea screening performance was improved with raw RRI data when a DL-based method is adopted for model training.

The purpose of this study is to develop a high-performance and real-driving-implementable driver drowsiness detection method using raw RRI data and DL technologies. The proposed method uses long short-term memory (LSTM) and autoencoder (AE) as DL architectures. LSTM (Gers et al., 2000) is a type of recurrent neural network (RNN) whose hidden layers include three gates that control long-term memory: an input gate, an output gate, and a forget gate. These gates respectively read, write, and forget memories, which contribute to learning long-term dependencies of time series data. AE is a type of neural network trained so that its outputs become close to its inputs, which is a typical DL architecture in the anomaly detection field (Sakurada and Yairi, 2014). In the proposed method, LSTM was embedded in AE whose inputs are raw RRI time series.

In this study, the drowsiness detection performance of the proposed method is evaluated through experiments using a driving simulator (DS).

2. METHOD & MATERIALS

2.1 Network Architecture

The network architecture adopted in the proposed method consists of a 30 length-input, a 30 length-output that duplicates the input, and hidden layers, including LSTM networks, which was referred to as LSTM-AE. A schematic diagram of LSTM-AE is shown in Fig. 1. First, an input one-dimensional time series data is expanded to a multi-dimensional time series by the first fully connected layer (FC). Next, the time series features are extracted in hidden layers with LSTM and FC. The blocks colored with green

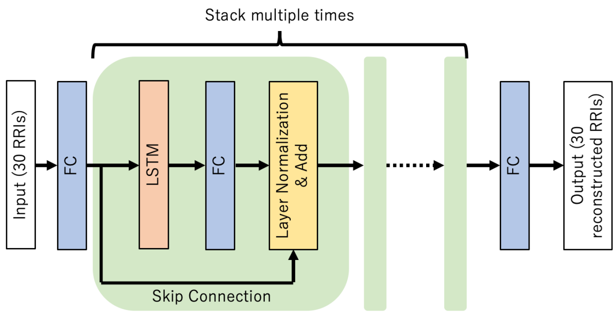


Fig. 1. Schematic diagram of LSTM-AE

in Fig. 1 are stacked multiple times. Each block has a skip connection that is expected to promote training (He et al., 2016). The output of the last FC is a time series with the same length as inputs.

Since the inputs and the outputs on LSTM-AE have the same dimension, LSTM-AE can be trained so that the output sequence becomes close to the input sequence. It is expected that a reconstruction error (RE) between the input sequence and the output sequence becomes large when the characteristics of the input sequence are significantly different from the training data. Thus, the abnormal sequence can be detected by monitoring RE. In this study, RE was defined as the mean squared error (MSE) between the input sequence $\{x_1, \ldots, x_n\}$ and the output sequence $\{\hat{x}_1, \ldots, \hat{x}_n\}$:

$$RE = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \tag{1}$$

where n is the length of the input sequence.

2.2 Drowsiness Detection Algorithm

Before using the proposed method, a drowsiness detection model needs to be trained from RRI data collected during awake driving. It is difficult to collect a large amount of drowsy driving RRI data for training and to collect awake driving RRI data is much easier than drowsy driving RRI data collection.

The training procedure is described in Algorithm 1. In this algorithm, $\boldsymbol{y}_n$ denotes nth RRI time series in the training data ($n = 1, 2, \cdots, N$). In steps 1-4, each $\boldsymbol{y}_n$ is rearranged to one matrix $\boldsymbol{Y}_n$ whose each row is a segmented RRI data with a length of thirty, and they are merged into one matrix $\boldsymbol{Y}$. The merged matrix $\boldsymbol{Y}$ is normalized to zero mean and unit variance in step 5. Finally, the drowsiness detection model with the LSTM-AE architecture is trained from $\tilde{\boldsymbol{Y}}$ in step 6.

Algorithm 1 Drowsiness detection model training

Require: A set of RRI time series $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N\}$.
1: **for** i = 1, . . . , N **do**
2: Rearrange $\boldsymbol{y}_i$ to $\boldsymbol{Y}_n$.
3: **end for**
4: Merge $\boldsymbol{Y}_n$ $(n = 1, \ldots, N)$ into one matrix $\boldsymbol{Y}$.
5: Normalize $\boldsymbol{Y}$, which is referred to as $\tilde{\boldsymbol{Y}}$.
6: Train a drowsiness detection model f from $\tilde{\boldsymbol{Y}}$.
7: **return** f.

Algorithm 2 Drowsiness detection

Require: The trained drowsiness detection model f.

1: Set the estimated driver state $S = \mathcal{A}$.
2: Set the time count $\tau = 0$.
3: **for** $i = 1,\ 2,\ \cdots$ **do**
4:   Measure a newly observed RRI x_i.
5:   **if** $i \geq 30$ **then**
6:     Construct the ith input sequence $\boldsymbol{y}_i = [x_{i-29},\ x_{i-28},\ \cdots,\ x_i]$.
7:     Normalize $\boldsymbol{y}_i$, which is referred to as $\tilde{\boldsymbol{y}}_i$.
8:     Calculate the ith output sequence $\hat{\boldsymbol{y}}_i$ from $\tilde{\boldsymbol{y}}_i$ using f.
9:     Calculate the ith RE e_i between $\tilde{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{y}}_i$.
10:     **if** $((e_i > \bar{e}) \wedge (S = \mathcal{A})) \vee ((e_i \leq \bar{e}) \wedge (S = \mathcal{D}))$ **then**
11:       $\tau \Leftarrow \tau + x_i$
12:     **else**
13:       $\tau \Leftarrow 0$
14:     **end if**
15:     **if** $\tau \geq \bar{\tau}$ **then**
16:       $S \Leftarrow \neg S$
17:       $\tau \Leftarrow 0$
18:     **end if**
19:   **end if**
20:   Wait until the next RRI is measured.
21: **end for**

The procedure of the proposed drowsiness detection is described in Algorithm 2. The estimated driver state is denoted as the binary variable $S \in \{\mathcal{A},\ \mathcal{D}\}$ which represents "awake" or "drowsy". That is, $\neg\mathcal{A} = \mathcal{D}$ and vice versa. When the ith RRI x_i $(i \geq 30)$ is observed in step 4, the ith input sequence $\boldsymbol{y}_i \in \mathbb{R}^{30}$ is constructed and normalized in steps 5-6. In steps 7-8, the ith output sequence $\hat{\boldsymbol{y}}_i$ is calculated from $\tilde{\boldsymbol{y}}_i$ using the drowsiness detection model f, and the ith RE e_i is calculated from $\tilde{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{y}}_i$.

If e_i exceeds its predefined threshold $\bar{e}$ for more than the predefined period $\bar{\tau}$, the estimated driver state S changes from $\mathcal{A}$ to $\mathcal{D}$. Conversely, when e_i continuously stays below $\bar{e}$ for more than $\bar{\tau}$, S changes from $\mathcal{D}$ to $\mathcal{A}$. Here, $\bar{e}$ and $\bar{\tau}$ are parameters for tuning the drowsiness detection performance.

2.3 Driving Simulator Experiment

In order to obtain the training data and evaluate the proposed method, a driving simulation experiment was performed. In the experiment, the participants drove a simulated vehicle along a monotonous course for three hours in a dark room. ECG, EEG, and electrooculogram (EOG) during driving were measured. The EEG and EOG electrode allocation is shown in Fig. 2; seven electrodes for EEG and two for EOG. ECG, EEG, and EOG were measured with the Grapevine Neural Interface Processor system (Ripple Neuro inc., USA), whose sampling frequency was 1,000 Hz. The used driving simulator was the UC-win/road system (FORUM8 Co., Ltd.), consisting of a steering wheel, a driving seat, and three LCD monitors displaying the front scene. The Research Ethics Committee of the Graduate School of Informatics, Kyoto University approved the experiment. All participants checked and signed informed consent prior to the experiment.

Table 1. Data profile

	awake [h]	drowsy [h]	total [h]
training	12.9	–	12.9
hyperparameter tuning	5.3	1.3	6.6
evaluation	12.1	2.1	14.2
total	30.3	3.4	33.7

Thirty-one university students having valid driving licenses participated in the experiment. The data of six participants were excluded because of measurement failure or inappropriate body conditions. The remaining 25 participants' data (male: 17, female: 8, 21±1.8 years old) were used for analysis.

2.4 Drowsiness Detection Preparation

The RRI data were extracted from the collected ECG data using a first derivative-based peak detection algorithm. Based on the collected EEG and EOG data, a sleep specialist certificated by the Japanese Society of Sleep Research scored sleep-related events such as microsleeps and slow eye movements (SEM). These events increase the risk of accidents (Moller et al., 2006; Shin et al., 2011). Based on these event presences, the RRI data were split into two-minute epochs and labeled as drowsy or awake. Since EEG scoring results may alter by sleep specialists, the data containing ambiguous scoring were excluded. Details of the RRI data used in the analysis are shown in Table 1, which shows that there were 12.9 hours of awake data for training, 6.6 hours of data (awake: 5.3 hours, drowsy: 1.3 hours) for hyperparameter tuning for LSTM-AE, and 14.2 hours of data (awake: 12.1 hours, drowsy: 2.1 hours) for evaluation of the proposed method. The tuned parameters in LSTM-AE are the number of stacking blocks, the number of LSTM units, the dimensions of the FC inputs in the stacking blocks and the final FC inputs, and the number of training epochs. In this study, the time threshold $\bar{\tau}$ was fixed at 10 seconds, and the RE threshold $\bar{e}$ was determined to maximize the difference between the true positive rate (TP rate) and the false positive rate (FP rate) of the hyperparameter tuning data. Here, TP means successful drowsiness detection, and FP means misdetection of awake states.

3. RESULT

In this study, the proposed method was evaluated by each two-minute epoch. Figure 3 shows RE examples of

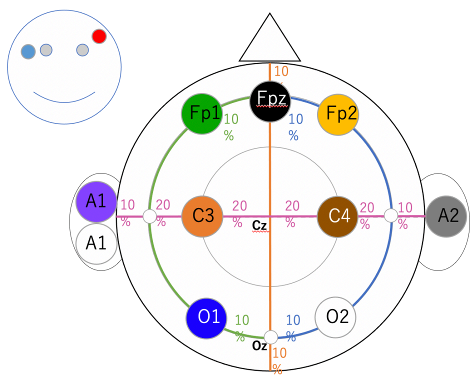
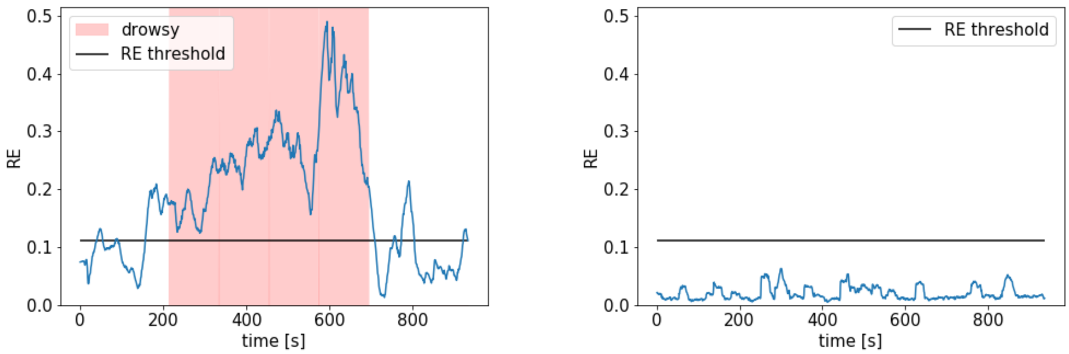


Fig. 2. Electrode allocation for EEG and EOG measurement



(a) Drowsy epochs (Participant 24)

(b) Awake epochs (Participant 4)

Fig. 3. REs of drowsy epochs and awake epochs

drowsy epochs and awake epochs: the RE of eight epochs containing four drowsy epochs in Fig. 3(a), in which four drowsy epochs are colored with red. In addition, the RE of eight epochs with awake epochs is shown in Fig. 3(b). These figures clearly illustrate that RE became highly high in the drowsy epochs.

The overall performance of the drowsiness detection by the proposed method is shown as a Receiver Operating Characteristic curve (ROC curve) in Fig. 4; the horizontal axis and the vertical axis represent FP rate and TP rate, respectively. In addition, the Area Under the ROC curve (AUC) was 0.88. The overall TP rate and FP rate were 81.3 % and 9.1 %, respectively. These results show that the proposed method detected drowsiness with a low FP rate.

4. DISCUSSION

The drowsiness detection result for each participant is shown in Table 2, whose columns are participant numbers, lengths of RRI data used for LSTM-AE training, the numbers of evaluation epochs, TP rates, and FP rates. Table 2 shows that FPs rarely occurred except in specifics participants. The trained drowsiness detection model seemed to be sufficiently generalized since participants 2 and 6, whose RRI data were not used for training, had no FPs.

The FP rates were significantly high in participants 24 and 30. According to the sleep scoring results, their awake epochs were close to drowsy epochs, which were within at most 10 minutes before or after drowsy epochs. The RE might increase around drowsy epochs as well as in drowsy epochs. This indicates that there is the possibility that the proposed method can be used to detect drowsiness prediction as well as detection.

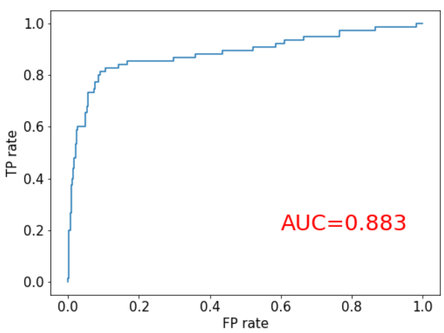


Fig. 4. ROC curve

Extra two methods were applied to the same dataset for comparison. One is a method combining HRV analysis and MSPC proposed by Fujiwara et al. (2019), and the other is almost the same as the proposed method except for input that uses eight HRV features instead of the raw RRI data. This method is referred to as HRV-LSTM-AE. The HRV features used in these two methods are shown in Appendix, and their comparison results are summarized in Table 3.

According to Table 3, the AUC of the proposed method was the highest of the three methods, and HRV-LSTM-AE the second. Their difference was the machine learning algorithms used for training. Although MSPC uses PCA that is a linear method, HRV-LSTM-AE uses DL, which is a highly nonlinear method. This fact suggests that it is possible to improve performance using a nonlinear method because it may handle complex physiological phenomena.

The difference between the proposed method and HRV-LSTM-AE is the used input data. Thus, the result indicates that the raw RRI data might be more appropriate than HRV features for driver drowsiness detection, which agrees with the result shown by Iwasaki et al. (2021). The conventional HRV feature extraction may lose some information related to sleep conditions.

In addition, the proposed method may be more robust against RRI artifacts than the HRV-based methods. It requires an RRI time series with a length of 30 as inputs shorter than HRV feature extraction, usually for three-minute. This means that the influence of RRI artifacts in the proposed method is shorter than the HRV-based methods, which may reduce FPs in the proposed method.

5. CONCLUSION

In this study, the new drowsiness detection method based on LSTM-AE was proposed. The experimental result using the driving simulator showed that the proposed method achieved an AUC of 0.88, a sensitivity of 81%, and a specificity of 91%, which were higher than methods using HRV features. This indicates that the use of the raw RRI

Table 2. Drowsiness detection result for each participant

No.	training data [h]	evaluation epochs	TP rate	FP rate
2	–	10	–	0 / 10
3	–	15	0 / 2	0 / 13
4	0.55	29	–	0 / 29
5	–	15	–	2 / 15
6	–	17	–	0 / 17
7	0.93	22	0 / 1	0 / 21
8	0.11	13	–	2 / 13
9	0.27	31	0 / 8	0 / 23
10	0.33	17	–	0 / 17
11	0.98	15	–	0 / 15
12	0.11	8	7 / 8	–
13	–	18	18 / 18	–
14	0.38	10	–	0 / 10
17	0.30	9	–	0 / 9
18	1.76	16	–	1 / 16
21	1.29	8	–	0 / 8
22	–	9	–	5 / 9
23	–	6	0 / 2	0 / 4
24	–	31	20 / 20	11 / 11
25	1.73	16	–	0 / 16
27	0.67	35	–	0 / 35
28	1.15	14	–	4 / 14
29	1.06	15	–	0 / 15
30	–	22	16 / 16	5 / 6
31	1.06	16	–	1 / 16
total	12.93	417	61 / 75	31 / 342

data may be more appropriate than HRV analysis for driver drowsiness detection.

There are some limitations in this study; the experimental environment was the driving simulator. Since the participants were not at the risk of traffic accidents, they might not seriously fight against drowsiness during driving. In addition, the participants were young because they were recruited from students at Kyoto University. Since HRV depends on ages (Shaffer and Ginsberg, 2017), the validation with aged participants will be needed.

In future work, additional experimental data must be collected to improve the drowsiness detection performance, and the system under development will be tested in a real driving environment.

Appendix A. HRV FEATURES USED FOR COMPARISON

The standard eight HRV features were used for comparison, which are classified into time-domain features and frequency-domain features. The time-domain features are obtained from the original RRI data.

- **meanNN :** Mean of RRI.
- **SDNN :** Standard deviation of RRI.
- **Total Power (TP) :** Variance of RRI.
- **RMSSD :** Root means square of the difference of adjacent RRI.
- **NN50 :** Number of pairs of adjacent RRI, whose difference is more than 50 ms.

Table 3. Comparison results

method	Proposed	Fujiwara	HRV-LSTM-AE
input	raw RRIs	HRV features	HRV features
Machine learning	LSTM-AE	PCA	LSTM-AE
AUC	0.88	0.72	0.80

Frequency-domain features are defined based on the power spectrum density (PSD) of the resampled RRI data.

- **LF :** Power of the low-frequency band (0.04Hz – 0.15Hz) in PSD. LF reflects the activity of both the sympathetic and parasympathetic nervous systems.
- **HF :** Power of the high-frequency band (0.15Hz – 0.4Hz) in PSD. HF reflects the parasympathetic nervous system activity.
- **LF/HF :** The ratio of LF to HF. LF/HF expresses the balance between the sympathetic nervous system activity and the parasympathetic nervous system activity.

In HRV feature extraction, a rectangular moving window was used, and the window size was three minutes, and the window moves every RRI measurement.

REFERENCES

Arefnezhad, S., Samiee, S., Eichberger, A., and Nahvi, A. (2019). Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection. *Sensors (Switzerland)*, 19(4). doi: 10.3390/s19040943.

Chiou, C.Y., Wang, W.C., Lu, S.C., Huang, C.R., Chung, P.C., and Lai, Y.Y. (2020). Driver Monitoring Using Sparse Representation with Part-Based Temporal Face Descriptors. *IEEE Transactions on Intelligent Transportation Systems*, 21(1), 346–361. doi: 10.1109/TITS.2019.2892155.

Fujiwara, K., Abe, E., Kamata, K., Nakayama, C., Suzuki, Y., Yamakawa, T., Hiraoka, T., Kano, M., Sumi, Y., Masuda, F., Matsuo, M., and Kadotani, H. (2019). Heart Rate Variability-Based Driver Drowsiness Detection and Its Validation With EEG. *IEEE Transactions on Biomedical Engineering*, 66(6), 1769–1778. doi: 10.1109/TBME.2018.2879346.

Gers, F.A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471. doi:10.1162/089976600300015015.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., and Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences. *Journal of Sleep Research*, 15(1), 47–53. doi:10.1111/j.1365-2869.2006.00504.x.

Iwasaki, A., Nakayama, C., Fujiwara, K., Sumi, Y., Matsuo, M., Kano, M., and Kadotani, H. (2021). Screening of sleep apnea based on heart rate variability and long short-term memory. *Sleep and Breathing*. doi:10.1007/s11325-020-02249-0.

Lin, C.T., Chang, C.J., Lin, B.S., Hung, S.H., Chao, C.F., and Wang, I.J. (2010). A real-time wireless brain–computer interface system for drowsiness detection. *IEEE transactions on biomedical circuits and systems*, 4(4), 214–222.

Malik, M., Bigger, J.T., Camm, A.J., Kleiger, R.E., Malliani, A., Moss, A.J., and Schwartz, P.J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European heart journal*, 17(3), 354–381.

Moller, H.J., Kayumov, L., Bulmash, E.L., Nhan, J., and Shapiro, C.M. (2006). Simulator performance, microsleep episodes, and subjective sleepiness: normative data using convergent methodologies to assess driver drowsiness. *Journal of Psychosomatic Research*, 61(3), 335–342. doi:10.1016/j.jpsychores.2006.04.007.

Rahman, A., Sirshar, M., and Khan, A. (2015). Real time drowsiness detection using eye blink monitoring. In *2015 National Software Engineering Conference (NSEC)*, 1–7. doi:10.1109/NSEC.2015.7396336.

Sakurada, M. and Yairi, T. (2014). Anomaly detection using autoencoders with nonlinear dimensionality reduction. *ACM International Conference Proceeding Series*, 02-December-2014, 4–11. doi:10.1145/2689746.2689747.

Shaffer, F. and Ginsberg, J.P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, 5(September), 1–17. doi:10.3389/fpubh.2017.00258.

Shin, D., Sakai, H., and Uchiyama, Y. (2011). Slow eye movement detection can prevent sleep-related accidents effectively in a simulated driving task. *Journal of Sleep Research*, 20(3), 416–424. doi:10.1111/j.1365-2869.2010.00891.x.

Wu, T., Redouté, J.M., and Yuce, M. (2019). A wearable, low-power, real-time ecg monitor for smart t-shirt and iot healthcare applications. In G. Fortino and Z. Wang (eds.), *Advances in Body Area Networks I*, 165–173. Springer International Publishing, Cham.

Zhang, C., Wu, X., Zheng, X., and Yu, S. (2019). Driver drowsiness detection using multi-channel second order blind identifications. *IEEE Access*, 7, 11829–11843. doi:10.1109/ACCESS.2019.2891971.