Elsevier Editorial System(tm) for Applied Ergonomics

Manuscript Draft

Manuscript Number:

Title: Classification of a Driver's Cognitive Workload Levels using Artificial Neural Network on ECG Signals

Article Type: Full Length Article

Keywords: Cognitive workload classification, heart rate variability, artificial neural network

Corresponding Author: Dr. Kihyo Jung, Ph.D.

Corresponding Author's Institution: University of Ulsan

First Author: Amir Tjolleng

Order of Authors: Amir Tjolleng; Kihyo Jung, Ph.D.; Wongi Hong; Wonsup Lee; Baekhee Lee; Heecheon You; Joonwoo Son; Seikwon Park

Abstract: An artificial neural network (ANN) model was developed in the present study to classify the level of a driver's cognitive workload based on electrocardiography (ECG). ECG signals were measured on 15 male participants while they performed a simulated driving task as a primary task with/without an N-back task as a secondary task. Three time-domain ECG measures (mean inter-beat interval (IBI), standard deviation of IBIs, and root mean squared difference of adjacent IBIs) and three frequency-domain ECG measures (power in low frequency, power in high frequency, and ratio of power in low and high frequencies) were calculated. To compensate for individual differences in heart response during the driving tasks, a three-step data processing procedure was performed to ECG signals of each participant: (1) selection of two most sensitive ECG measures, (2) definition of three (low, medium, and high) cognitive workload levels, and (3) normalization of the selected ECG measures. An ANN model was constructed using a feed-forward network and scaled conjugate gradient as a back-propagation learning rule. The accuracy of the ANN classification model was found satisfactory for learning data (95%) and testing data (82%).

Cover Letter

February 11, 2016

Subject: Submission of new manuscript for peer review

Dear Editor-in-Chief,

I am enclosing herewith a manuscript entitled “Classification of a Driver’s Cognitive Workload Levels using Artificial Neural Network on ECG Signals”. The manuscript has not been published in any other journals.

Sincerely yours,

Kihyo Jung

Corresponding author: Kihyo Jung, Ph. D.

School of Industrial Engineering,
University of Ulsan,
93 Daehak-ro, Nam-gu, Ulsan 680-749, Korea
Tel: +82-52-259-2709 Fax: +82-52-259-1683
E-mail: kjung@ulsan.ac.kr

Highlights

- An artificial neural network (ANN) model was developed to classify the level of cognitive workload.
- A three-step data processing was performed to compensate for individual differences in heart response.
- Six ECG measures in time (mean IBI, SDNN, and RMSSD) and frequency (LF, HF, and LF/HF) domains were collected.
- Accuracy of the ANN model was found satisfactory for learning data (95%) and testing data (82%).

1. Introduction

Cognitive workload and drowsiness during driving are considered major causes of vehicle accidents. The National Safety Council (NSC) reported that cognitive workload causes 28% of all crashes (NSC, 2010). The National Highway Traffic Safety Administration estimated that 100,000 accidents per year in the USA were caused by driver drowsiness (Rau, 2005). Additionally, Yamakoshi et al. (2008), Eoh et al. (2005), Aidman et al. (2015), Pack et al. (1995) and Williamson et al. (2011) reported that driver overload and monotony are two significant causative factors in traffic accidents. Therefore, the detection of cognitive workload and drowsiness during driving is important for preventing accidents and hazards on the road (Engström et al., 2005; Verwey and Zaidel, 1999; Wong and Huang, 2009).

The physiological responses of drivers have been widely used in the detection of cognitive workload and drowsiness in a vehicle. Eoh et al. (2005) and Lin et al. (2005) observed a significant drop in the alpha of electroencephalogram (EEG) as drowsiness increased. Mayser et al. (2003) and Jagannath & Balasubramanian (2014) found a decrease in electromyogram (EMG) as cognitive workload and drowsiness increased. Genno et al. (1997), Ohsuga et al. (2001), and Yamakoshi et al. (2008) observed a decrease in skin temperature with increased cognitive workload and drowsiness. Lastly, Milosevic (1997), Yang et al. (2010), and Patel et al. (2011) found a decrease in mean inter-beat interval (IBI) of electrocardiograph (ECG) with increased cognitive workload and an increase in mean IBI with increased drowsiness.

Among the aforementioned physiological responses, ECG is considered a reliable measure in estimating a driver's status. ECG signals can be quantified in terms of time and frequency domains. Time domain measures include mean IBI, standard deviation of IBIs (SDNN), and root mean squared difference of adjacent IBIs (RMSSD) (Combatalade, 2010; Juan, 2004). These time domain measures decrease when the level of cognitive workload increases (Berntson et al., 1997; Brookhuis and Waard, 2001; Mehler et al., 2009; Wood et al., 2002). Frequency domain measures include power in low frequency (LF), power in high frequency (HF), and LF/HF ratio (Calcagnini et al., 1994; Tal and

David, 2000; Yang et al., 2010). The LF and LF/HF ratio increase and the HF decreases as the level of cognitive workload increases (Wood et al., 2002).

On the other hand, there is an individual variation in heart response. Many studies have reported that heart responses to tasks show significant differences between individuals (Hong et al., 2014; Lee et al., 2010; Lal and Craig, 2001). First, an effective ECG measure varies noticeably among individuals. For example, the RMSSD of Driver A in Figure 1.a changes more by cognitive tasks than other ECG measures, while the mean IBI of Driver B in Figure 1.b is more distinctly altered than the other measures. Next, heart sensitivity to cognitive tasks of different levels varies individually. For example, a low workload task for Driver A in Figure 1.a can be differentiated from medium and high workload tasks, while a high workload task for Driver B in Figure 1.b can be discriminated from low and medium workload tasks. Lastly, the magnitudes of ECG measures also vary among individuals. For example, Driver A in Figure 1.a shows a smaller mean IBI than Driver B in Figure 1.b for all cognitive tasks.

[Insert Figure 1 about Here]

Although advanced classification methods have been applied in the detection of drowsiness and cognitive workload, the classification accuracy for cognitive workload needs to be improved. Patel et al. (2011) used an artificial neural network (ANN) to identify the presence of driver drowsiness and reported a classification accuracy of 90%. In addition, Vicente et al. (2011) utilized a linear discriminant analysis to classify a driver into two statuses (awake or drowsy) and presented a specificity of 93% and a sensitivity of 85%. On the other hand, Zhang et al. (2014) applied a regression method to classify the extent of cognitive workload into two levels (normal or elevated workload) and showed an accuracy of 62.5%. Solovey et al. (2014) used five classification methods (decision tree method, logistic regression method, multilayer perceptron method, Naïve Bayes method, and nearest neighbor method) to classify the extent of workload into the two levels and reported an accuracy of 71.5% to 74.1%. Although several classification methods have been applied to classify the extent of cognitive workload level, their accuracies are low because they do not consider the individual differences of heart response by cognitive workload in the development of a classification model.

The present study developed an ANN model considering individual differences in classifying the level of a driver's cognitive workload based on ECG data. ECG data were measured while participants performed a simulated driving task as the primary task with/without an N-back task as the secondary task. The individual differences in heart response were adjusted at the signal processing stage. The ANN model was trained using a feed-forward network and back-propagation learning rule and then evaluated in terms of sensitivity and specificity.

2. Method and Materials

2.1. Participants

Fifteen male participants with at least 3 years of driving experience were recruited in this study. Their average (SD) age was 27.7 (3.0) and the participants were healthy and had no discomfort on the day of experiment. Their participation were compensated.

2.2. Equipment

A driving simulator (STISIM Drive™, Systems Technology Inc.) was used in this study, as shown in Figure 2. The driving simulator consisted of a vehicle and a large screen (resolution: 1024 × 768) showing a driving scene. The driving scenario was to drive on a two-lane (width of a lane: 4.57 m) highway at a speed of about 100 km/h.

[Insert Figure 2 about Here]

An ECG system (MEDAC System/3, Biomation) was used to measure ECG signals while the participants drove a driving simulator. Three ECG sensors were attached below the left clavicle, right clavicle, and left rib. The sampling rate was set to 250 Hz.

2.3. Experimental Design

The participants were instructed to drive (primary task) the driving simulator while performing an N-back task (secondary task). The N-back task was to recall the N step's earlier number when an experiment instructor presented a sequence of arbitrary numbers (Hong et al., 2014; Son et al., 2010). The level of difficulty of the N-back task could be adjusted based on N. Four driving tasks (driving without N-back task, driving with 0-back task, driving with 1-back task, and driving with 2-back task) were performed to simulate multitasking with different levels of difficulty.

The experiment was conducted in four steps. The purpose of the experiment was explained to the participant and informed consent was obtained. Next, ECG sensors were attached to the participant, and practice driving was allowed to be familiarized with the simulator driving and N-back tasks. Then, the main experiment was conducted and ECG data were collected during the four driving tasks lasting 2 minutes each. Lastly, a debriefing session was conducted.

2.4. Signal Processing

Measurements for six ECG measures in time (mean IBI, SDNN, and RMSSD) and frequency (LF, HF, and LF/HF) domains were collected in four steps. First, IBI data were calculated from the raw ECG signals using the R-peak detection algorithm (Billauer, 2012) coded in Matlab (MathWorks, Inc., USA). Second, the IBI data measured between 10 and 110 sec were selected for further analysis. Third, the three time domain measures were quantified using Equation 1, 2, and 3, respectively. Lastly, three frequency domain measures were obtained by fast Fourier transformation in Matlab. For frequency analysis, this study determined the appropriate time window to be 100 sec based on

Clifford (2002). The frequency bands for LF (0.04 - 0.15 Hz) and HF (0.15 - 0.4 Hz) were defined based on Combatalade (2010).

$$\text{Mean IBI} = \frac{\sum_{i=1}^{n} IBI_i}{n} \qquad (1)$$

where: n = number of inter-beat intervals,

IBI_i= i^{th} inter-beat interval

$$\text{SDNN} = \sqrt{\frac{\sum_{i=1}^{n}(IBI_i - \overline{IBI})}{n-1}} \qquad (2)$$

where: n = number of inter-beat intervals,

IBI_i= i^{th} inter-beat interval,

$\overline{IBI}$= average of inter-beat intervals

$$\text{RMSSD} = \sqrt{\frac{\sum_{i=1}^{n-1}(IBI_{i+1} - IBI_i)^2}{n-1}} \qquad (3)$$

where: n = number of inter-beat intervals,

IBI_i= i^{th} inter-beat interval

To adjust for individual differences in heart response, the following three-step signal processing procedure was conducted: (1) selection of two sensitive ECG measures, (2) definition of three workload levels, and (3) normalization of the selected ECG measures. In the first step, the two sensitive ECG measures for each participant were selected from the six ECG measures. Since the sensitivities of the ECG measures were different among participants, two ECG measures which best discriminated the driving tasks were selected for each participant. For example, in Figure 3.a, mean IBI and RMSSD were selected as sensitive measures due to their systematic trend of change with different driving tasks.

[Insert Figure 3 about Here]

In the second step, the three workload levels were individually defined for each participant by grouping the four driving tasks. Since the level of perceived workload based on the driving tasks varied among participants, the four driving tasks of each participant were grouped into three workload categories (low, medium, and high). For example in Figure 3.b, a participant was less sensitively changed during the driving and driving with 0-back tasks than during other driving tasks. Thus, the participant's perceived workload level was defined as low (driving and driving with 0-back tasks), medium (driving with 1-back task), and high (driving with 2-back task).

In the last step, the two selected ECG measures were normalized by their medians. The magnitude of the ECG measures varied significantly among participants. To eliminate this individual difference, the values of the selected ECG measures were normalized using each individual participant's median value. Figure 3.c illustrates the normalizing process for a participant using Equation 4.

$$\mathrm{N}_i = \frac{x_i}{\tilde{x}} \qquad (4)$$

where: $N_i = i^{\text{th}}$ normalized data

$x_i = i^{\text{th}}$ data

$\tilde{x}$ = median

2.5. ANN Modeling

The topology of the ANN model consisted of three layers (input, hidden, and output layers) as shown in Figure 4. The input layer had two nodes for the two normalized ECG measures. The hidden layer, which processed the normalized ECG measures using the sigmoid activation function, had 15 neurons. The number of neurons in the hidden layer affected the classification accuracy; however, no accepted theory currently exists for predetermining the optimal number of neurons (Acharya et al., 2003). Hence, the optimal number of neurons (15) was determined by varying it from 5 to 30 with an interval of 5 until the network with highest sensitivity and specificity was obtained. The output layer had three nodes, which denoted three levels (low, medium, and high) of cognitive workload.

[Insert Figure 4 about Here]

A standard feed-forward and back-propagation neural network was employed in the present study. A three layer feed-forward network was utilized in the Neural Network Toolbox of Matlab. A hyperbolic tangent sigmoid transfer function was applied as the transfer function of the hidden layer. A linear transfer function was used for the transfer function of the output layer. The scaled conjugate gradient was utilized as a back-propagation network learning function. Lastly, the ECG data of the fifteen participants were randomly divided into learning and testing sets--70% of the ECG data for learning of the ANN model and the remaining for testing.

3. Results

3.1. ECG Measures

The time domain measures were more sensitive to changes in workload than frequency domain measures as shown in Figure 5. The time domain measures (normalized mean IBI, SDNN, and RMSSD) gradually declined as the workload level increased. For example, the normalized mean IBI was 1.05 (0.80 sec) for the low workload, 1.00 (0.77 sec) for the medium workload, and 0.94 (0.72 sec) for the high workload. Meanwhile, the frequency domain measures (normalized LF, HF, and LF/HF ratio) showed insignificant changes with change in workload. For example, the normalized LF was 0.99 (0.1107 m^2) for the low workload, 1.00 (0.1117 m^2) for the medium workload, and 1.01 (0.1137 m^2) for the high workload.

[Insert Figure 5 about Here]

A one-factor (workload level) within-subject ANOVA test of the six normalized ECG measures revealed that the normalized mean IBI ($F(2, 28) = 17.58$, $p < 0.001$) and normalized RMSSD ($F(2, 28) = 9.84$, $p = 0.001$) were significantly altered by the workload level at $\alpha = 0.05$. Tukey tests classified the workload levels into three groups (Group A: low workload, Group B: medium workload, and Group C: high workload) for the normalized mean IBI and two groups (Group A: low workload, Group B: medium and high workload) for the normalized RMSSD. On the other hand, the normalized SDNN and the three frequency measures showed a systematic trend with the elevation of cognitive workload, but it was not statistically significant (normalized SDNN: $F(2, 28) = 1.64$, $p = 0.212$; normalized LF: $F(2, 28) = 1.84$, $p = 0.178$; normalized HF: $F(2, 28) = 0.91$, $p = 0.414$; normalized LF/HF: $F(2, 28) = 2.42$, $p = 0.108$).

3.2. ANN Performance

The classification accuracy of the proposed ANN was satisfactory for both the learning and testing sets. The cross evaluation was repeated 100 times to rigorously validate the performance of the proposed ANN. The average classification accuracies for the learning and testing sets were 95% (SD = 2.77) and 82% (SD = 8.58), respectively. As shown in Figure 6, sensitivity (true positive rate) and specificity (true negative rate) had no systematic bias in the learning and testing sets.

[Insert Figure 6 about Here]

4. Discussion

An ANN model considering individual differences in heart responses was developed to accurately classify the level of drivers' cognitive workload based on ECG data. Two sensitive ECG measures of each participant were selected to correct the individual difference in the sensitivity of ECG measures. Three levels (low, medium, and high) of cognitive workloads were defined for each participant by grouping four driving tasks (driving without N-back task, driving with 0-back task, driving with 1-

back task, and driving with 2-back task) to adjust the individual difference in the perceived extent of workload. In addition, the ECG measures were normalized by its median to correct the individual difference in the magnitude of ECG signal. The ANN model developed in this study showed high classification accuracies for both the learning (95%) and testing (82%) data sets. The ANN model can be applied to the development of an intelligent vehicle which identifies elevated cognitive workload and provides biofeedback to prevent a vehicle accident.

Mean IBI decreased gradually to come up with an oxygen demand as the workload level increased. The normalized mean IBI in this study was 1.05 for the low workload, 1.00 for the medium workload, and 0.94 for the high workload, which can be explained by the relationship between cognitive workload and oxygen demand. A cognitive overload promotes oxygen demand by cells and leads to the production of more cardiac output by increasing heart rate (Brookhuis et al., 1991; Brookhuis and Waard, 2001; Lenneman et al., 2005; Mehler et al., 2009). Since heart rate is inversely proportional to mean IBI (heart rate = 60 sec / mean IBI), cognitive overload decreases mean IBI.

SDNN and RMSSD also decreased with an increase in the level of cognitive workload, which can be explained by the role of the sympathetic nerve and the parasympathetic nerve in the autonomic nervous system. Under a high cognitive workload, the sympathetic nerve is activated and stabilizes heart rate to more stably produce cardiac outputs (Low, 2013; Camm et al., 1996). Otherwise, under a low workload, the parasympathetic nerve takes this role, which leads to a fluctuation in heart rate.

LF and HF changed in the opposite way as the level of cognitive workload increased. Since LF was dominantly affected by the sympathetic nerve (Billman, 2013; Bezerianos et al., 1999), a high cognitive workload increased LF by activating the sympathetic nerve. On the other hand, HF was mainly influenced by the parasympathetic nerve; thus, a low cognitive workload increased HF by activating the parasympathetic nerve.

Cognitive workload influenced ECG measures differently from drowsiness. The mean IBI decreased as cognitive workload increased, while the mean IBI increased as drowsiness increased (Lal and Craig, 2001). In addition, the calculated LF/HF ratio in this study increased when the difficulty of the workload increased; on the other hand, the LF/HF ratio significantly decreased with drowsiness

(Patel et al., 2011). Thus, it is implied that cognitive workload and drowsiness modulate the sympathetic and parasympathetic nerves in an opposite manner.

Future research is needed to enhance the applicability of the proposed ANN model in the development of an intelligent vehicle. First, an in-depth evaluation for various age and gender groups is required to comprehensively understand the relationship between cognitive workload and ECG measures. The present study only recruited young male drivers in the experiment. Since age and gender affect the sensitivity of heart response (Mehler et al., 2009), participants of varying age and gender are necessary for generalization of the present study results. Lastly, a field study is needed in a real vehicle to validate the results of the present study because the experiment in the present study was conducted in a driving simulator in which driving conditions and environment were controlled.

Acknowledgments

This study was supported by the Research Fund of the University of Ulsan.

References

Acharya, U. R., Bhat, P. S., Iyengar, S. S., Rao, A., and Dua, S. (2003). Classification of heart rate data using artificial neural network and fuzzy equivalence relation. *Pattern Recognition*, 36, 61 – 68.

Aidman, E., Chadunow, C., Johnson, K., and Reece, J. (2015). Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. *Accident Analysis and Prevention*, 81, 8–13.

Berntson, G. G., Bigger Jr, J. T., Eckberg, D. L., Grossman,P., Kaufmann, P. G., Malik, M., Nagaraja, H. N., Porges, S. W., Saul, J. P., Stone, P. H., and Van DerMolen., M. W. (1997). Heart rate variability: origins, methods, and interpretive caveats. *Psychophysiology*, 34, 623-648.

Bezerianos, A., Papadimitriou, S., and Alexopoulos, D. (1999). Radial basis function neural networks for the characterization of heart rate variability dynamics. Artificial Intelligence in Medicine, 15, 215–234.

Billauer, E. (2012). Peak detection using MATLAB. Retrieved July 20, 2012 from http://www.billauer.co.il/peakdet.html.

Billman, G. E. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in Physiology*, 4, 1-5.

Brookhuis, K. A., and De Waard, D. (2001). Assessment of drivers' workload: performance, subjective and physiological indices. In Hancock, P. and Desmond P.(Eds.). *Stress, Workload and Fatigue: Theory, Research and Practice*. New Jersey: Lawrence Erlbaum.

Brookhuis, K. A., G. De Vries, and D. De Waard (1991). The effects of mobile telephoning on driving performance. *Accident Analysis and Prevention*, 24 (3), 309-316.

Calcagnini, G., Biancalana, G., Giubilei, F., Strano, S., and Ceruti, S. (1994). Spectral analysis of heart rate variability signal during sleep stages. *Proceedings of the 16th Annual IEEE International Conference*, pp. 1252–1253.

Camm, A. J., Bigger, J. T. Jr, Breithardt, G., Cerutti, S., Cohen, R. J., Coumel, P., Fallen, E. L., Kennedy, H. L., Kleiger, R. E., Lombardi, F., Malliani, A., Moss, A. J., Rottman, J. N., Schmidt, G., Chewartz, P. J., and Singer, D., H. (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17, 354-381.

Clifford, G. D. (2002). *Signal processing Methods for Heart Rate Variability*. Unpublished Doctoral Thesis. University of Oxford.

Combatalade, D. C. (2010). Basics of Heart Rate Variability Applied to Psychophysiology. Thought Technology Ltd., Canada.

Engström, J., Johansson, E., and Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research F*, 2, 97–120.

Eoh, H. J., Chung, M. K., and Kim, S-H. (2005). Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics*, 35, 307–320.

Genno, H., Ishikawa, K., Kanbara, O., Kikumoto, M., Fujiwara, Y., Suzuki, R., and Osumi, M. (1997).Using facial skin temperature to objectively evaluate sensations. *International Journal of Industrial Ergonomics*, 19, 161-171.

Hong, W., Lee, W., Jung, K., Lee, B., Park, J., Park, S., Park, Y., Son, J., Park, S., and You, H. (2014). Development of an ECG-based assessment method for a driver's cognitive workload. *Journal of the Korean Institute of Industrial Engineers*, 40(3), 325-332.

Jagannath, M. and Balasubramanian, B. (2014). Assessment of early onset of driver fatigue using multimodal fatigue measures in a static simulator. *Applied Ergonomics*, 45, 1140-1147.

Juan, S. (2004). Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. *Swiss Med Weekly*, 134, 514-522.

Lal, S.K.L., and Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, 55, 173-194.

Lee, W., Jung, K., Hong, W., Park, S., Park, Y., Son, J., Park, S., and Kim, K. (2010). Analysis of drivers' ECG biological signal under different levels of cognitive workload for intelligent vehicle. *Proceedings of the 2010 Fall Conference of Ergonomics Society of Korea.*

Lenneman, J.K., Shelley, J.R., and Backs, R.W. (2005). Deciphering psychological-physiological mappings while driving and performing a secondary memory task. *Proceedings of the Third International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*.

Lin, C. T., Wu, R. C., Liang, S. F., Chao, W. H., Chen, Y. J., and Jung, T.P. (2005). EEG-based drowsiness estimation for safety driving using independent component analysis. *IEEE Transactions on Circuits and Systems I : Regular Papers*, 12 (52).

Low, P. (2013). Overview of the Autonomic Nervous System. Home Health Handbook. Retrieved October 10, 2015 from http://www.merckmanuals.com.

Mayser, C., Piechulla, W., Weiss, K.-E., and König, W. (2003). Driver workload monitoring. In H. Strasser, K. Kluth, H. Rausch, & H. Bubb (Eds.).Quality of Work and Products in Enterprises of the Future. *Proceedings of the 50th Anniversary Conference of the GfA and the XVII Annual ISOES Conference in Munich*.

Mehler, B., Reimer, B., Coughlin, J. F., and Dusek, J. A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Proceedings of Transportation Research Board 88th Annual Meeting*.

Milosevic, S. (1997). Drivers' fatigue studies. *Ergonomics*, 40(3), 381-389.

National Safety Council (NSC). (2010). Understanding the distracted brain: Why driving while using hands-free cell phones is risky behavior.

Ohsuga, M., Shimono, F., and Genno, H. (2001). Assessment of phasic work stress using autonomic indices. *International Journal of Psychophysiology*, 40, 211-220

Pack, A.I., Pack, A.M., Rodgman, E., Cucchiara, A., Dinges, D.F., and Schwab, C. W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis & Prevention*, 27(6), 769–775.

Patel, M., Lal, S.K.L., Kavanagh, D., and Rossiter, P. (2011). Applying neural network analysis on heart rate variability data to assess driver fatigue. *Expert Systems with Applications*, 38, 7235–7242.

Rau, P.S. (2005). Drowsy driver detection and warning system for commercial vehicle drivers: Field proportional test design, analysis, and progress. National Highway Traffic Safety Administration.

Solovey, E. T., Zee, M., Perez, E., Reimer, B., and Mehler, B. (2014). Classifying driver workload using physiological and driving performance data: two field studies. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 4057-4066

Son, J., Reimer, B., Mehler, B., Pohlmeyer, A. E., Godfrey, K.M., Orszulak, J., Long, J., Kim, M.H., Lee, Y. T., and Coughlin, J. F. (2010). Age and cross-cultural comparison of drivers' cognitive workload and performance in simulated urban driving. *International Journal of Automotive Technology*, 11(4), 533-539.

Tal, O. G., and David, S. (2000). Driver fatigue among military truck drivers. *Transportation Research Part F*, 3, 195–209.

Verwey, W. B., and Zaidel, D. M. (1999). Preventing drowsiness accidents by an alertness maintenance device. *Accident Analysis and Prevention*, 31, 199–211.

Williamson, A., Lombardi, D.A., Folkard, S., Stutts, J., Courtney, T.K., and Connor, J. L. (2011). The link between fatigue and safety. *Accident Analysis and Prevention*, 43, 498–515.

Wong, J.T. and Huang, S.H., (2009). Modelling driver mental workload for accident causation and prevention. *Journal of the Eastern Asia Society for Transportation Studies*, 8, 1918-1933.

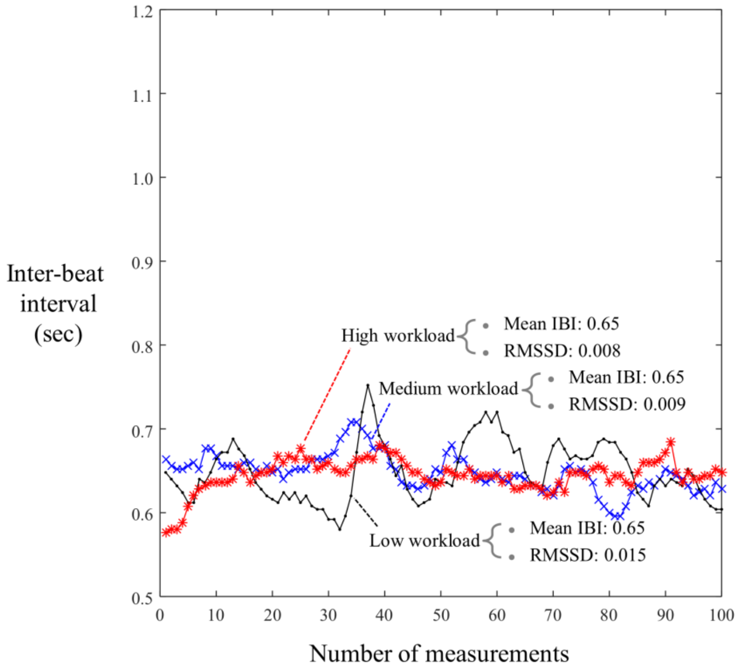
Wood, R., Maraj, B., Lee, C. M., and Reyes, R. (2002). Short-term heart rate variability during a cognitive challenge in young and older adults. *Age and Aging*, 31, 131-135.

Yamakoshi, T., Yamakoshi, K., Tanaka, S., Nogawa, M., Park, S. B., Shibata, M., Sawada, Y., Rolfe, P., and Hirose, Y. (2008). Feasibility study on driver's stress detection from differential skin temperature measurement. *Proceedings of the 30th Annual International IEEE EMBS Conference*, 1076–1079.

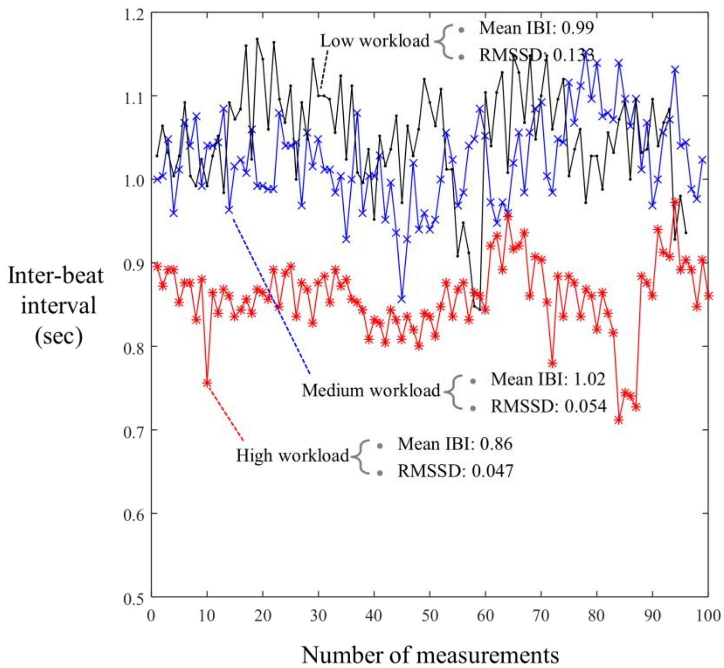
Yang, G., Lin, Y., and Bhattacharya, P. (2010). A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences*, 180, 1942-1954.

Zhang, H., Zhu, Y., Maniyeri, J., and Guan, C. (2014). Detection of variations in cognitive workload using multi-modality physiological sensors and a large margin unbiased regression machine. *Engineering in Medicine and Biology Society, 2014 36th Annual International Conference of the IEEE*, 2985-2988.

List of Figures



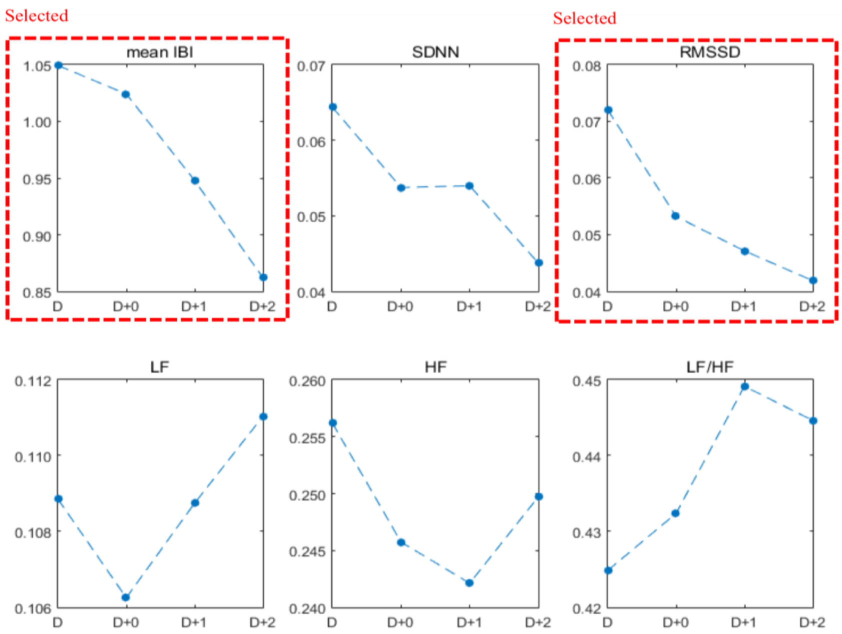
(a) Driver A



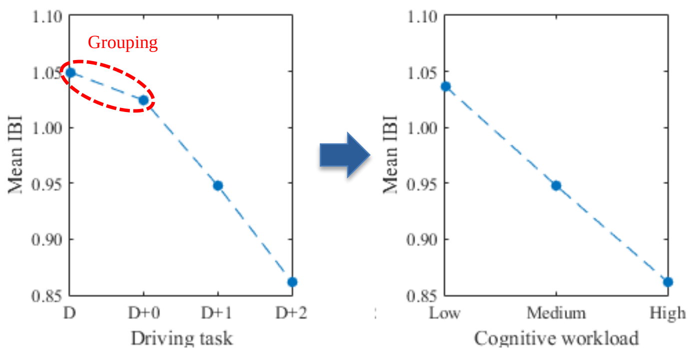
(b) Driver B

Figure 1. Illustration of ECG changes based on cognitive workload

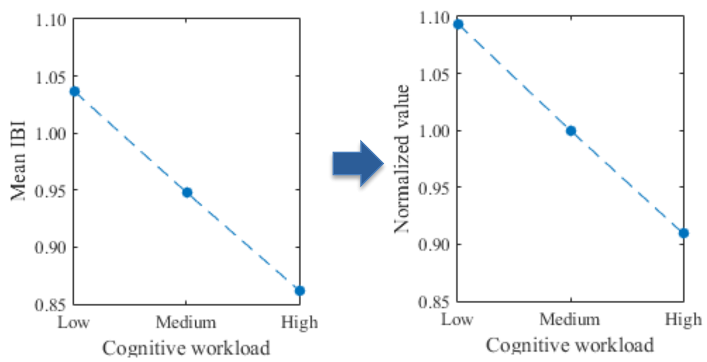
Figure 2. Driving simulator used in this study



(a) Selection of two sensitive ECG measures



(b) Definition of the three workload levels based on driving tasks



(c) Normalization of the ECG measure

Figure 3. Illustration of correction for individual differences (D: driving, D+0: driving with 0 back task, D+1: driving with 1 back task, D+2: driving with 2 back task)

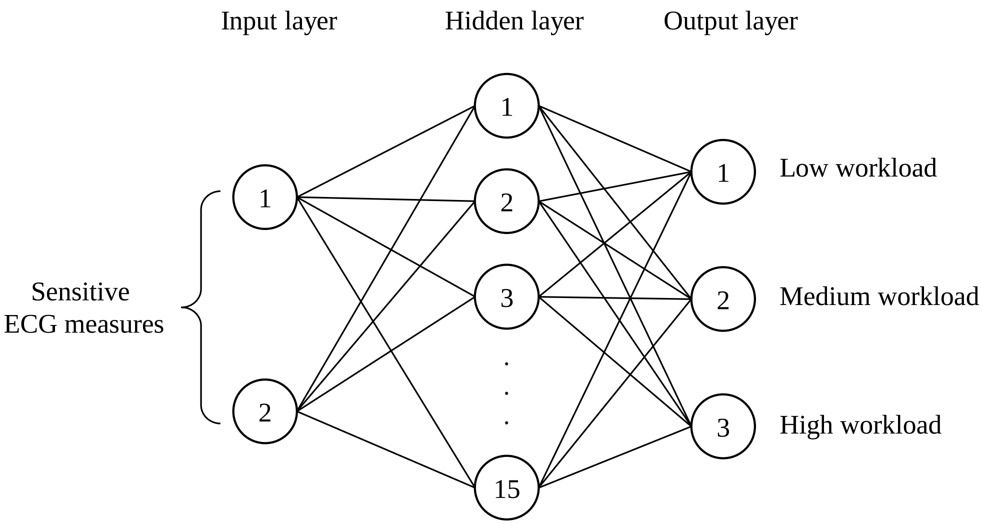


Figure 4. Three-layer feed-forward neural network structure

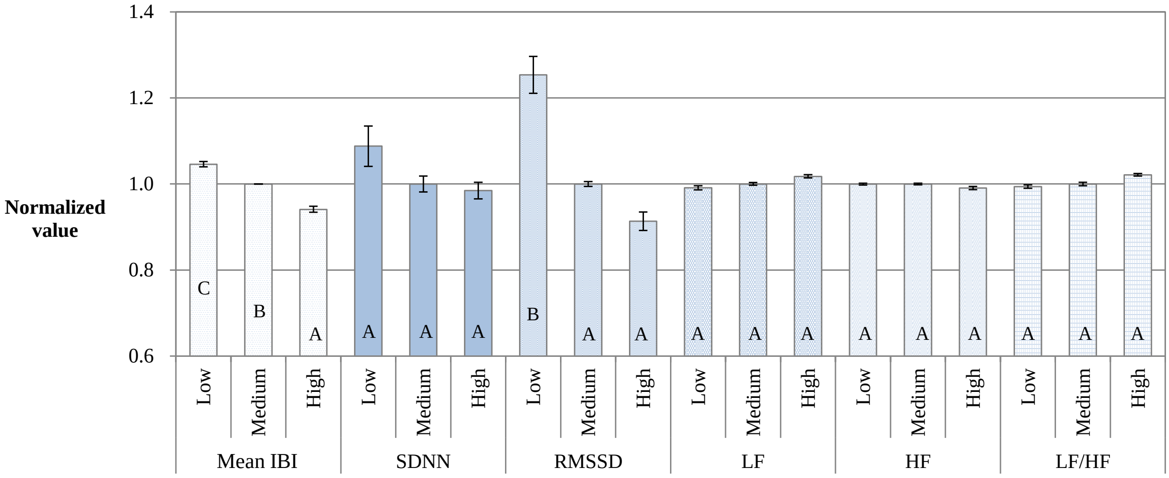
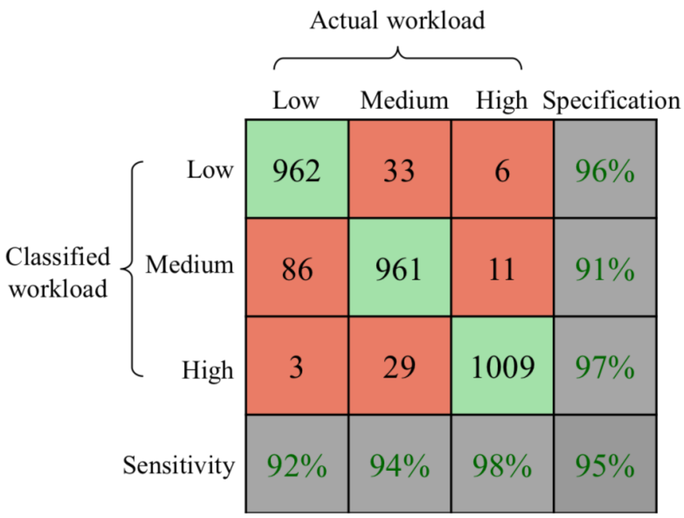
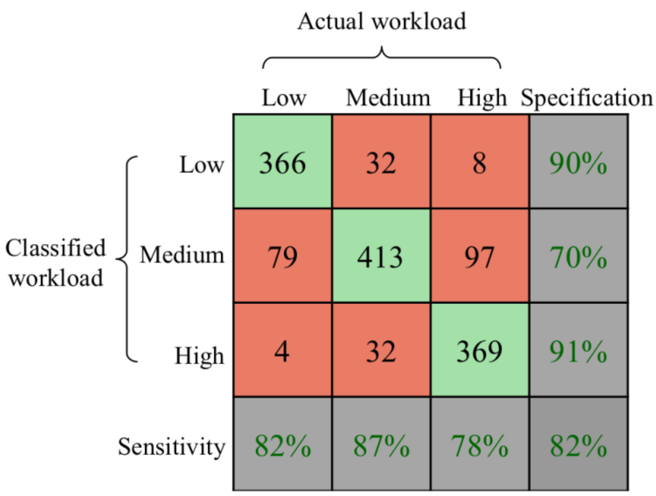


Figure 5. Normalized ECG measures for different workloads

(Note: Different alphabet letters indicate statistically significant differences at the 0.05 level)



(a) Learning data set



(b) Testing data set

Figure 6. Confusion matrix

(Note: the diagonal cells in each matrix show the number of cases that were correctly classified, and the off-diagonal cells show the misclassified cases. The bottom cells show sensitivity and the right cells display specificity. The bottom right cell in each matrix shows the total percent of correctly classified cases.)