Cognitive Workload Detection from Raw EEG-Signals of Vehicle Driver using Deep Learning

Mohammad A. Almogbel*, Anh H. Dang**, Wataru Kameyama***

*Department. of Computer Science and Communications Engineering,
Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo, Japan.

**GITS, Waseda University, Tokyo, Japan

***Faculty of Science and Engineering, Waseda University, Tokyo, Japan

almogbela@ruri.waseda.jp, danghoanganh@akane.waseda.jp, wataru@waseda.jp

***Abstract*—Electroencephalography (EEG) signals have been proven to be effective in evaluating human's cognitive state under specific tasks. Conventional classification models utilized for EEG classification heavily rely on signal pre-processing and hand-designed features. In this paper, we propose an end-to-end deep neural network which is capable of classifying multiple types of cognitive workload of a vehicle driver and the context of driving using only raw EEG signals as its input without any pre-processing nor the need for conventional hand-designed features. Data used in this study are collected throughout multiple driving sessions conducted on a high-fidelity driving simulator. Experimental results conducted on 4 channels of raw EEG data show that the proposed model is capable of accurately detecting the cognitive workload of a driver and the context of driving.**

***Keywords*— Deep Learning, EEG, Neural Networks, Cognitive Workload, Driving, Stress**

I. INTRODUCTION

Maintaining safe levels of cognitive workload is extremely crucial to ensure optimal performance and attention whilst driving automobiles. Different methods have been widely investigated to monitor driver's cognitive workload including heart rate monitoring [1][2], galvanic skin response [3][4], facial expression [5][6] and so on [7][8][9][10]. One of the most popular measures for assessing the mental state of humans is electroencephalography (EEG) signals. EEG has reportedly shown success in evaluating the mental state such as drowsiness [11], mind wandering [12] and alertness [13].

Despite being an active research area, understanding and applying EEG signals are still limited due to the lack of true understanding of the brain's activities which prevents use of good EEG features from being engineered. Besides, EEG signals are usually strongly affected by noise and interference. Accurate reading requires a delicate equipment and sealing.

Conventional analyzing approaches make extensive use of Fourier transform in order to decompose EEG signal into multiple frequencies. Only well-known frequency is then used for feature design process. For example, the alpha band (8 to 15[Hz]) is correlated with relaxing, and the beta band (16 to 31[Hz]) associates with mental stress. Pre-processing methods such as Butterworth bandpass filter and stationary wavelet transform filters [14] are used to remove high and low frequency noises [13]. Several efforts in applying deep learning on EEG have been carried out [11][19][20][21]. However, all the previous researches still rely on signal pre-processing pipelines, and none come up with a system that can handle raw data directly. Thus, valuable information might be discarded during the pre-processing.

The purpose of this paper is to introduce an end-to-end deep neural network model that can directly infer cognitive workload and context of driving from raw EEG, where signal pre-processing as well as conventional hand-designed features are not required.

To evaluate the proposed model, EEG signal recordings have been carried out on a subject driving a vehicle in a relatively realistic simulated environment under several different types of cognitive workloads and contexts of driving.

Our experiment shows that the proposed model can accurately classify different labels of cognitive workload and the driving context from raw EEG signals of vehicle driver in a simulated environment without the use of any conventional hand-designed features.

This paper is organized as follows: Section II previews the related researches. Section III explains the proposed model used to classify the data. Section IV details the data collection methodology, and the conducted experiments for evaluating the proposed model. Section V summarizes the acquired results, and finally the conclusion is given in Section VI.

II. RELATED WORKS

Monitoring electrical brain activity in a non-invasive means is referred to as EEG in clinical context. Brain activities produce electrical charges caused by the neurons inside the brain, and voltage fluctuations are measured using conductive material called electrodes and a reference electrode attached to the head and scalp [15]. These voltages pass through an amplifier for analysis. Thanks to recent advancements, compact lightweight devices such as MindWave Mobile Headset [16], Emotiv Insight 5 [17] and Muse [18] which is used in this study, have been introduced

Manuscript received on Dec 31, 2017.
Mohammad A. Almogbel is with Graduate School of Fundamental Science and Engineering, Waseda University, Tokyo, Japan (Corresponding author, email: almogbela@ruri.waseda.jp)
Anh H. Dang is with GITS, Waseda University, Tokyo, Japan. (email: danghoanganh@akane.waseda.jp)
Wataru Kameyama is with Faculty of Science and Engineering, Waseda University, Tokyo, Japan (email: wataru@waseda.jp)

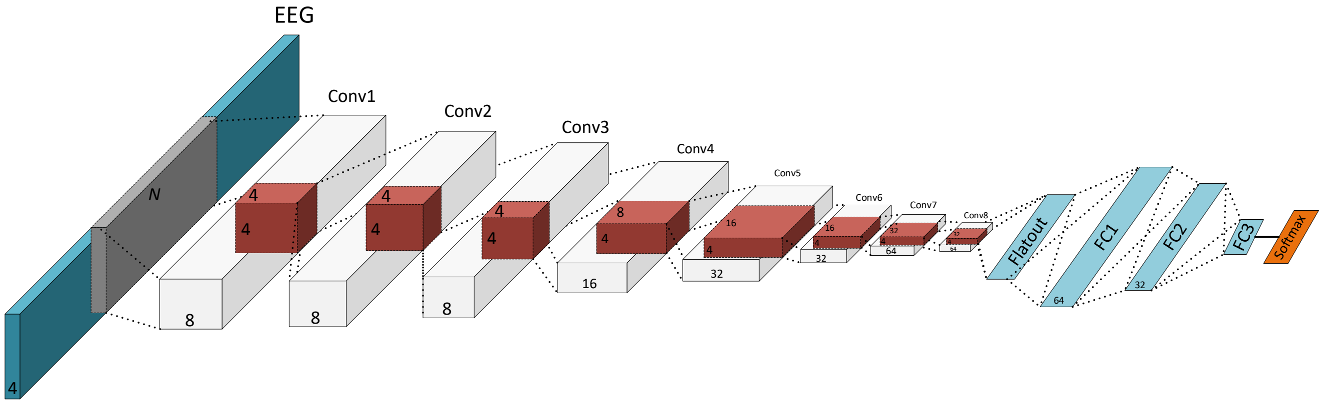

Fig. 1 Network Diagram. The blue cuboid denotes the four raw EEG channels while the gray cuboid inside denotes the slicing window. The white cuboid denotes the kernel while the red cubes denotes the filters applied within each convolutional layer. All the configuration are detailed in Table II.

into the market to allow a more consumer friendly means to monitor EEG signals.

Conventional approaches to classify EEG signals consist of decomposing the signals to extract features to be used for classification. Recent studies of EEG analysis to utilize deep learning with different pre-processing methodologies are shown in Table I.

In [11], FFT is used before classifying the signals to detect driver's drowsiness. [19] evaluates drivers' cognitive performance in a simulated environment using filtered frequency EEG. Similarly, [20] also proposes a model that can predict right and left hand movements with frequency filtered EEG signals. [21] applies a spatial filter before classifying pathological from normal EEG recordings.

TABLE I
RECENT STUDIES ON EEG ANALYSIS WITH DEEP LEARNING

Ref.	Pre-processing	Feature-Extraction	Classifier	Accuracy
[19]	Frequency filtering	CNN	1-layer ANN	86.06%
[11]	FFT	N/A	1-layer ANN	86.50%
[21]	Spatial Filter	CNN	1-layer ANN	84.80%
[20]	Frequency filtering	CNN	1-layer ANN	86.41%

III. PROPOSED MODEL

One of the major advantages of deep neural networks is the ability to extract, learn and generalize features without pre-processing [22]. In this study, we propose the use of deep convolutional networks to predict the cognitive workload of a driver and the context of driving by using raw EEG signals with no additional pre-processing.

CNN is inspired by biological visual system [23] where different levels of visual hierarchy are handled by different parts of the brain. Analogically, multi-layers of CNN generalize visual features in multiple levels. Since each layer of CNN acts as a filter, it is expected that CNN can be also used to extract high level features even in the case of EEG signals.

The architecture of the proposed network is visualized in Fig. 1. The network takes a vector of $4 \times N$ as its input where 4 is the number of raw EEG channels collected from Muse device and N is the length of the signal. Throughout the paper we test the network with different lengths which is explained in detail in section IV-A.

The model consists of several convolutional layers that transform raw EEG data to feature vectors. These features are then flattened and processed by 3 layers of fully connected (FC) classifier blocks.

The FC block outputs a two-, three- or six-dimensional vector depending on the type of classification. Softmax function is used to interpret this vector into the probability of each classification class. Except the final layer of FC block, all layers in this deep network utilize rectified linear unit (RELU) [24] activation function. Batch normalization [25] is also used after every RELU activation function.

Our CNN filters are applied to all EEG channels in the same time. This method ensures that all possible combinations of features are captured. Small filter size and stride together with the high number of CNN layers also ensure that features from multiple hierarchical levels are extracted.

Different from images, EEG comprises multiple time series from each channel. For that reason, it is essential to design CNN filters to stride along time direction only. Thus, it must fully cover all other dimensions of input data.

Table II describes the network configuration for the proposed model. The output shape of each layer is omitted from the table since it is tested on different input sizes. Input sizes are explained in Table III in Section IV-A.

TABLE II
NETWORK CONFIGURATION

Convolutional Block			
	Filter Size	Strides	Activation
Conv1	8×4×4	2	RELU → BN
Conv2	8×4×4	2	RELU → BN
Conv3	8×4×4	2	RELU → BN
Conv4	16×4×8	2	RELU → BN
Conv5	32×4×16	2	RELU → BN
Conv6	32×4×16	2	RELU → BN
Conv7	64×4×32	4	RELU → BN
Conv8	64×4×32	4	RELU → BN
Flat-out			
Fully-Connected Block			
	No. units	Dropout	Activation
FC1	64	0.5	RELU → BN
FC2	32	0.5	RELU → BN
FC3	6/3/2	-	-

IV. EXPERIMENT

A. Data Collection

1) EEG

EEG recording has been done using Muse [18] headband. Fig. 2 shows the international 10-20 system that is used to describe the location of the electrodes on the head. The figure also explains the electrodes used in this device where 4-

channel configuration has been utilized; two dry electrodes on the forehead AF7 (Left) and AF8 (Right), two behind the ears TP9 (Left) and TP10 (Right), and a reference electrode (Fpz) in the middle of the forehead above the nasion. Muse is capable of recording EEG signals at the sampling rate of 256[Hz]. Moreover, it is very portable, lightweight, easy to use, and it utilizes low-energy Bluetooth technology to transmit data making it feasible for real-life usages.

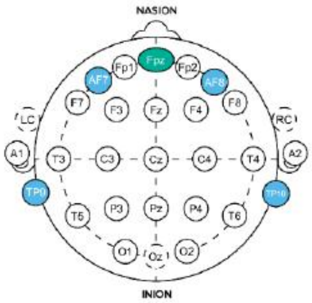

Fig 2. Locations of Scalp Electrodes in the International 10-20 System [18]

Raw EEG data are normalized by z-score for each of the 4 channels. Slicing window approach is utilized to prepare the data for training, where each slicing window is denoted by the time segments of the EEG data. Table III explains the different slicing window size used for training the model as well as the shape of the input data for the network. The slicing windows are overlapped with 1/256[sec] step.

TABLE III
INPUT SHAPES

Raw EEG Slicing Window Size	Input Shape
256Hz × 180 sec × 4 channels	46080 × 4
256Hz × 150 sec × 4 channels	38400 × 4
256Hz × 120 sec × 4 channels	30720 × 4
256Hz × 90 sec × 4 channels	23040 × 4
256Hz × 60 sec × 4 channels	15360 × 4
256Hz × 30 sec × 4 channels	7680 × 4

2) Driving Simulator

Data used in this study are collected through recordings conducted on a popular driving simulation video game (GTA-V free driving mode) [26] that is highly capable of representing the real life driving situations. GTA-V has achieved an increasing popularity amongst recent studies that utilizes driving simulators [27][28][29][30]. Other used equipment consists of Logitech G27 driving wheel and an ultrawide (21:9) curved monitor to better simulate natural vision inside a real car. The experiment environment is shown in Fig. 3.

Fig 3. Representation of the used Simulation Environment with Examples of City (Left) and Highway (Right).

3) Driving Sessions

The driving sessions were designed to accommodate broad scenarios of multiple cognitive workloads and contexts of driving. The data collection was conducted by breaking down the traffic flow/density inside the simulator environment into three levels; zero traffic (light), moderately dense traffic (medium) and high dense traffic (high). Table IV shows the traffic flow and density for each level where traffic flow denotes the average number of cars passing a single point per hour and traffic density denotes the average number of cars within 1 kilometer of a single point.

TABLE IV
EXPERIMENT CONDITIONS AND COMBINATIONS

	Traffic		Type Combinations	
	Flow	Density	City	Highway
Light	0 cars/min	0 cars/km	LC	LH
Medium	30 cars/min	22.5 cars/km	MC	MH
High	36 cars/min	33.23 cars/km	HC	HH

For each traffic flow/density, *city* and *highway* recordings were conducted, which means that the subject was advised to drive exclusively in both city and highway for each traffic flow/density. By doing so, six different types of driving scenarios are accumulated to evaluate the proposed model. Table IV also shows the experiment conditions and combinations with their abbreviations which are used hereafter, such as LC for light traffic in city.

Twelve sessions were collected for each of the six types, where one session was recorded for 15 minutes, which sums up to 72 sessions and 18 hours of raw EEG signals sampled at 256[Hz]. A total of 16,588,800 data points for each of the four raw channels was collected.

The recordings were conducted on a span of two months with random intervals and different times throughout the day. In this experiment, one subject of male with age 29 was joined, who drives frequently with 12 years of driving experience.

B. Types of Classification

To evaluate our proposed model, we considered the following classification problems:

1) Cognitive Workload Classification:

In this type of classification, we are interested in classifying the cognitive workload of the driver regardless of the driving context. In order to carry out this experiment, we group 6 types of driving sessions as explained in section IV-A into 3 classes with regards to the traffic flow/density,

- Low workload: LC, LH
- Medium workload: MC, MH
- High workload: HC, HH.

Traffic congestion is reportedly highly correlated with stress [31], and by grouping the sessions in such manner we investigate the correlation of EEG signals with the traffic/flow density.

Each class in this experiment consists of 24 merged sessions upon each traffic flow/density (5,529,600 × 4 data points). The last session of each merged class is used for evaluation. By doing so, the real-life situation, where the model requires to estimate the workload from future data by learning previous data, can be simulated.

Slight modifications in the network are as follows; the number of the units in FC1 is decreased to 32 in slicing window sizes of 90, 60 and 30 [sec]; it remains as the default configuration shown in Table II in slicing windows of size 180, 150 and 120 [sec].

2) Context Classification:

In this type of classification, we focus on classifying the

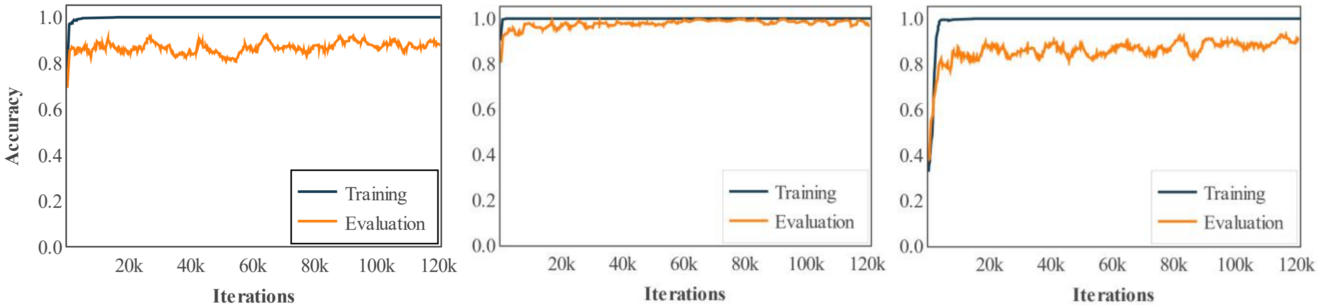

Fig. 4 Learning curve for training and evaluation of the model with the best performing slicing window size (150 [sec]) for workload (left), context (middle) and workload/context (right)

context of driving regardless of the cognitive workload of the driver. Inferring the context of driving directly from EEG signals can be helpful in analyzing the brain signals activity during different context of driving. Therefore, we group our EEG data into 2 classes: city driving and highway driving;

- City: LC, MC, HC
- Highway: LH, MH, HH

Each of the two classes yields 36 sessions (8,294,400 × 4 data points) for each of city and highway sessions. The last session of each merged class is used for evaluation. In this type of classification, the network configuration is unchanged from the default parameters as in Table II.

3) Context/Workload Classification:

Finally, we are interested in classifying the cognitive workload of the driver as well as the driving context at the same time. We assign a unique label to every cognitive workload and context mentioned above to perform a six-way classification for detecting both the cognitive workload of the driver and the context of driving.

Each of the six types yields 2,764,800 data points for every raw EEG channel. The last session for each type is used for evaluation. Slight modifications to the model are as follows; for the slicing window size of 180[sec], the number of units in FC1 is increased to 128. And for the slicing window sizes of 60/30 [sec], the number of units in the FC1 is decreased to 32 instead of 64, and the strides for the 7^{th} convolutional layer is lowered to a stride of 2 down from a stride of 4. The rest remains the same as the default configuration shown in Table II.

C. Training and Evaluation Parameters

The model is trained for 10 epochs using RMSProp optimizer with a learning rate of 0.0001. An epoch represents the training of the model throughout all the data. Input data is divided into batches of 64. During the training the process 50% dropout is applied to the first and second FC layers.

To evaluate the model, a random batch of 64 and its corresponding label is sampled from the evaluation set to be used for cross-validation, and for every batch; recall, precision and accuracy scores are calculated.

V. Results & Discussion

Fig. 4 shows the learning curve for training and evaluation of the model for the best performing slicing window size of 150[sec] for all three types of classification.

1) Workload Classification

We train our model with data prepared using different sizes of slicing windows. Table V shows the evaluation result of our proposed model for each window after the training process. The bold type denotes the highest score among different slicing window sizes. As shown in the table, the model achieves the highest performance with larger slicing window sizes. By average, 120[sec] is the optimal window sizes for this experiment.

Furthermore, the network is capable of accurately classifying *low* workload sessions, whereas *high* workload data are sometimes misclassified as *medium* workload. Session types are grouped together with regards to the traffic flow/density, however the actual cognitive workload could not be correlated in such manner in real life situations.

2) Context Classification

Similar to the workload classification, Table VI shows the evaluation results of our model on context classification. This type of classification achieves the best overall results, and by average, 150[sec] is the optimal window sizes for this

TABLE V
Workload Classification Metrics For different Window Sizes

		180	150	120	90	60	30	Avg.
Light	Recall	**0.997**	0.979	0.960	0.928	0.936	0.791	0.932
	Precision	**1.000**	0.996	1.000	0.982	0.987	0.982	0.991
	Accuracy	**0.999**	0.991	0.984	0.970	0.976	0.926	0.974
	F1 Score	**0.999**	0.987	0.980	0.954	0.961	0.876	0.959
Medium	Recall	0.833	0.921	0.905	0.915	0.890	**0.974**	0.906
	Precision	**0.997**	0.878	0.950	0.673	0.867	0.673	0.840
	Accuracy	0.950	0.923	**0.953**	0.834	0.912	0.838	0.902
	F1 Score	0.908	0.899	**0.927**	0.776	0.878	0.796	0.864
High	Recall	**1.000**	0.871	**1.000**	0.627	0.912	0.749	0.860
	Precision	0.884	0.926	0.900	0.912	0.900	**0.986**	0.918
	Accuracy	0.951	0.932	**0.969**	0.855	0.933	0.908	0.925
	F1 Score	0.938	0.898	**0.947**	0.743	0.906	0.851	0.881
Average	Recall	0.943	0.923	**0.955**	0.823	0.912	0.838	0.899
	Precision	**0.960**	0.933	0.950	0.856	0.918	0.880	0.916
	Accuracy	0.967	0.949	**0.969**	0.886	0.940	0.891	0.934
	F1 Score	0.948	0.928	**0.951**	0.824	0.915	0.841	0.901

TABLE VI
CONTEXT CLASSIFICATION METRICS FOR DIFFERENT WINDOW SIZES

		180	150	120	90	60	30	Avg.
City	Recall	0.885	**0.954**	0.862	0.918	0.886	0.903	0.901
	Precision	0.926	**0.939**	0.862	0.887	0.935	0.901	0.908
	Accuracy	0.996	0.998	0.985	**1.000**	0.995	0.981	0.976
	F1 Score	0.905	**0.947**	0.862	0.903	0.910	0.902	0.905
Highway	Recall	0.927	0.937	0.886	0.886	**0.946**	0.890	0.912
	Precision	0.890	**0.957**	0.886	0.921	0.901	0.896	0.908
	Accuracy	0.996	0.998	0.985	**1.000**	0.995	0.981	0.976
	F1 Score	0.908	**0.947**	0.886	0.903	0.923	0.893	0.910
Average	Recall	0.906	**0.946**	0.874	0.902	0.916	0.896	0.907
	Precision	0.908	**0.948**	0.874	0.904	0.918	0.898	0.908
	Accuracy	0.996	0.998	0.985	**1.000**	0.995	0.981	0.976
	F1 Score	0.906	**0.947**	0.874	0.903	0.916	0.897	0.907

experiment. The network can accurately distinguish between city and highway driving which shows a high correlation between the context of driving and the EEG signals.

3) Workload/Context Classification

Table VII shows the workload/context classification results. Similar to context classification, the window size of 150[sec] achieves the best performance by average. However, in general the network performance slightly drops in this type of classification. Since we train the network each class individually, each class yields smaller number of samples to train per class and the potential similarities between specific types of sessions such as MC and HH.

Low recall scores in LC and MH can be examined in smaller slicing window sizes due to the network misclassifying them for MC or HC.

Discussion

In Table I in section II we show recent reseach that utilize deep learning on EEG signals, however study does not impose in any way a direct comparison with the distinguished previous works because the used data, experimental conditions and classification targets are different in each, but rather explores and introduces the potential of using deep CNN architecture to eliminate the need for pre-processing and feature-extracion.

Nevertheless, our proposed model achieves high accuracy, recall and precision scores on raw EEG signals without applying any conventional pre-processing methodologies.

VI. CONCLUSION

Ever-proposed and conventional EEG analysis and classification studies are heavily reliant on signal pre-processing and hand-designed features. Such methodologies can be time consuming and potentially cause loss for viable information during the process. In this paper, we propose an end-to-end deep neural network which eliminates the necessity for such pre-processing pipelines, whilst achieving high classification performance.

Using only raw EEG signals from 4 channels as its input, the proposed model performs highly robust and accurate classification. The model is able to achieve an accuracy of 0.960 on average for the vehicle driver's cognitive workload and the context of driving.

Future works include testing the proposed model with public available data sets. Further investigation will be

TABLE VII
CONTEXT/WORKLOAD CLASSIFICATION METRICS FOR DIFFERENT WINDOW SIZES

`		180	150	120	90	60	30	Avg.
LC	Recall	0.745	0.773	0.744	**0.796**	0.685	0.704	0.748
	Precision	**1.000**	**1.000**	0.983	0.958	0.935	0.789	0.944
	Accuracy	0.963	**0.970**	0.951	0.958	0.942	0.911	0.949
	F1 Score	0.854	**0.872**	0.847	0.870	0.790	0.744	0.830
LH	Recall	**1.000**	0.994	0.972	0.889	0.995	0.711	0.928
	Precision	**1.000**	**1.000**	0.938	0.955	0.954	0.806	0.942
	Accuracy	**1.000**	**1.000**	0.984	0.977	0.992	0.926	0.980
	F1 Score	**1.000**	0.997	0.955	0.921	0.974	0.756	0.934
MC	Recall	0.722	0.682	0.693	0.892	0.679	**0.918**	0.757
	Precision	0.739	0.758	**0.771**	0.701	0.696	0.693	0.727
	Accuracy	0.898	0.905	0.905	0.913	0.895	**0.909**	0.904
	F1 Score	0.731	0.718	0.730	0.786	0.688	**0.789**	0.740
MH	Recall	0.725	**0.910**	0.809	0.898	0.517	0.807	0.777
	Precision	**1.000**	0.965	0.983	0.887	0.886	0.826	0.927
	Accuracy	0.955	**0.980**	0.973	0.960	0.914	0.945	0.954
	F1 Score	0.840	**0.937**	0.888	0.892	0.653	0.816	0.838
HC	Recall	0.932	0.918	**1.000**	0.635	0.867	0.784	0.857
	Precision	0.770	0.765	0.738	0.811	0.741	**0.914**	0.786
	Accuracy	0.939	0.927	0.928	0.913	0.919	**0.948**	0.929
	F1 Score	0.844	0.835	**0.849**	0.712	0.799	0.844	0.814
HH	Recall	0.932	0.918	**1.000**	0.635	0.867	0.784	0.951
	Precision	0.804	0.904	**0.921**	0.841	0.682	0.915	0.843
	Accuracy	0.954	0.977	**0.983**	0.953	0.905	0.964	0.956
	F1 Score	0.863	0.911	**0.959**	0.724	0.763	0.845	0.844
Avg.	Recall	0.843	0.866	**0.870**	0.791	0.768	0.785	0.836
	Precision	0.886	**0.899**	0.889	0.859	0.816	0.824	0.862
	Accuracy	0.952	**0.960**	0.954	0.946	0.928	0.934	0.945
	F1 Score	0.855	**0.878**	0.871	0.817	0.778	0.799	0.833

carried out by collecting more data from more subjects with different driving experiences.

REFERENCES

[1] N. Munla, M. Khalil, A. Shahin, and A. Mourad, “Driver stress level detection using HRV analysis,” *2015 Int. Conf. Adv. Biomed. Eng. ICABME 2015*, pp. 61–64, 2015.

[2] B. Eilebrecht *et al.*, “The relevance of HRV parameters for driver workload detection in real world driving,” *Comput. Cardiol. (2010).*, vol. 39, pp. 409–412, 2012.

[3] L. Boon-Leng, L. Dae-Seok, and L. Boon-Giin, “Mobile-based wearable-type of driver fatigue detection by GSR and EMG,” *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2016–Janua, pp. 1–4, 2016.

[4] V. Rajendra and O. Dehzangi, “Detection of distraction under naturalistic driving using Galvanic Skin Responses,” *2017 IEEE 14th Int. Conf. Wearable Implant. Body Sens. Networks, BSN 2017*, no. 2, pp. 157–160, 2017.

[5] M. Venturelli, G. Borghi, R. Vezzani, and R. Cucchiara, “Deep head pose estimation from depth data for in-car automotive applications,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10188 LNCS, pp. 74–85, 2018.

[6] H. Gao, A. Yuce, and J. P. Thiran, “Detecting emotional stress from facial expressions for driving safety,” *2014 IEEE Int. Conf. Image Process. ICIP 2014*, no. May 2015, pp. 5961–5965, 2014.

[7] A. Sahayadhas, K. Sundaraj, and M. Murugappan, “Detecting driver drowsiness based on sensors: A review,” *Sensors (Switzerland)*, vol. 12, no. 12, pp. 16937–16953, 2012.

[8] S. F. Liang, C. T. Lin, R. C. Wu, Y. C. Chen, T. Y. Huang, and T. P. Jung, “Monitoring driver’s alertness based on the driving performance estimation and the EEG power spectrum analysis.,” *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, vol. 6, pp. 5738–5741, 2005.

[9] A. Palazzi, D. Abati, S. Calderara, F. Solera, and R. Cucchiara, “Predicting the Driver’s Focus of Attention: the DR(eye)VE Project,” pp. 1–25, 2017.

[10] X. Li, H. Huang, and Y. Sun, “DriTri: An in-vehicle wireless sensor network platform for daily health monitoring,” *Proc. IEEE Sensors*, pp. 3–5, 2017.

[11] I. Belakhdar, W. Kaaniche, R. Djmel, and B. Ouni, “Detecting driver drowsiness based on single electroencephalography channel,” *13th Int. Multi-Conference Syst. Signals Devices, SSD 2016*, pp. 16–21, 2016.

[12] C. L. Baldwin, D. M. Roberts, D. Barragan, J. D. Lee, N. Lerner, and J. S. Higgins, “Detecting and Quantifying Mind Wandering during Simulated Driving,” *Front. Hum. Neurosci.*, vol. 11, no. August, pp. 1–15, 2017.

[13] L. Bi, R. Zhang, and Z. Chen, “Study on Real-time Detection of Alertness Based on EEG,” *2007 IEEE/ICME Int. Conf. Complex Med. Eng.*, pp. 1490–1493, 2007.

[14] S. S. Daud and R. Sudirman, “Butterworth Bandpass and Stationary Wavelet Transform Filter Comparison for Electroencephalography Signal,” *Proc. - Int. Conf. Intell. Syst. Model. Simulation, ISMS*, vol. 2015–Octob, pp. 123–126, 2015.

[15] S. Donald L. and F. H. L. da Silva, *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2011.

[16] NeuroSky, “MindWave Mobile Headset.,” 2015. [Online]. Available: https://store.neurosky.com/pages/mindwave.

[17] EMOTIV, “EMOTIV Insight 5 Channel Mobile EEG - Emotiv,” *Emotive*, 2018. [Online]. Available: https://www.emotiv.com/product/emotiv-insight-5-channel-mobile-eeg/.

[18] Interaxon, “Muse: the brain sensing headband,” *Tech. Specif. Valid. Res. use*, pp. 4–9, 2017.

[19] M. Hajinoroozi, Z. Mao, T. P. Jung, C. T. Lin, and Y. Huang, “EEG-based prediction of driver’s cognitive performance by deep convolutional neural network,” *Signal Process. Image Commun.*, vol. 47, pp. 549–555, 2016.

[20] Z. Tang, C. Li, and S. Sun, “Single-trial EEG classification of motor imagery using deep convolutional neural networks,” *Optik (Stuttg).*, vol. 130, pp. 11–18, 2017.

[21] R. T. Schirrmeister, L. Gemein, K. Eggensperger, F. Hutter, and T. Ball, “Deep learning with convolutional neural networks for decoding and visualization of EEG pathology,” 2017.

[22] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” *IEEE Int. Symp. Circuits Syst. Nano-Bio Circuit Fabr. Syst.*, pp. 253–256, 2010.

[23] K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, 1980.

[24] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” *Proc. 27th Int. Conf. Mach. Learn.*, no. 3, pp. 807–814, 2010.

[25] S. Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” Feb. 2015.

[26] “Gta V,” 2013. [Online]. Available: https://www.rockstargames.com/V/.

[27] Y. Chen, W. Li, and L. Van Gool, “ROAD: Reality Oriented Adaptation for Semantic Segmentation of Urban Scenes,” 2017.

[28] P. Bazilinskyy *et al.*, “Eye movements while cycling in GTA V,” no. January, pp. 7–11, 2018.

[29] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, “Driving in the Matrix: Can virtual worlds replace human-generated annotations for real world tasks?,” *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 746–753, 2017.

[30] M. Angus *et al.*, “Unlimited Road-scene Synthetic Annotation (URSA) Dataset,” 2018.

[31] D. A. Hennessy, D. L. Wiesenthal, and P. M. Kohn, “The Influence of Traffic Congestion, Daily Hassles, and Trait Stress Susceptibility on State Driver Stress: An Interactive Perspective 1,” 2000.

Mohammad A. Almogbel (S’14) received his bachelor’s degree in Information Systems from King Saud University, Riyadh, Saudi Arabia in 2009. He joined King Abdul-Aziz City for Science and Technology in Saudi Arabia as a researcher in the same year and received a scholarship to complete his graduate school in 2010. He then received master’s degree in computer science from Waseda University in 2014 and he continued to pursue his Ph.D. since then. He is a member of IEEE, ITS and JSAE.

Anh H. Dang (S’09) received his bachelor’s degree in business administration, information & communication technology from Ritsumeikan Asia Pacific University (Beppu, Oita, Japan) in 2010. He then received the master’s degree in computer science from Waseda University (Shinjuku, Tokyo, Japan) in 2012. Since 2012, he is a Ph.D. candidate at Waseda University. He is a member of IEEE, ACM, and IEICE. His research interests are machine learning, artificial intelligence, and computer vision.

Wataru Kameyama (M’86) received the bachelor’s, master’s, and D.Eng. degrees from the School of Science and Engineering, Waseda University, in 1985, 1987, and 1990, respectively. He joined ASCII Corporation in 1992 and was transferred to France Telecom CCETT from 1994 to 1996 for his secondment. After joining Waseda University as an Associate Professor in 1999, he has been a Professor with the Department of Communications and Computer Engineering, School of Fundamental Science and Engineering, Waseda University, since 2014. He has been involved in MPEG, MHEG, DAVIC, and the TV-Anytime Forum activities. He was a Chairman of ISO/IECTC1/SC29/WG12, and a Secretariat and Vice Chairman of the TV-Anytime Forum. He is a member of IEICE, IPSJ, ITE, IIEEJ, and ACM. He received the Best Paper Award of Niwa-Takayanagi in 2006, the Best Author Award of Niwa-Takayanagi in 2009 from the Institute of Image Information and Television Engineers, and the International Cooperation Award from the ITU Association of Japan in 2012.