Preprint Version. Manuscript submitted to and accepted by 2019 International Conference on Cyberworlds.

EEG-based Cross-subject Mental Fatigue Recognition

Yisi Liu
Fraunhofer Singapore
Singapore, Singapore
LiuYS@ntu.edu.sg

Zirui Lan
Fraunhofer Singapore
Nanyang Technological University
Singapore, Singapore
LanZ0001@e.ntu.edu.sg

Jian Cui
Fraunhofer Singapore
Nanyang Technological University
Singapore, Singapore
Cuijian@ntu.edu.sg

Olga Sourina
Fraunhofer Singapore
Nanyang Technological University
Singapore, Singapore
EOSourina@ntu.edu.sg

Wolfgang Müller-Wittig
Fraunhofer Singapore
Nanyang Technological University
Singapore, Singapore
Wolfgang.Mueller-wittig@fraunhofer.sg

Abstract—**Mental fatigue is common at work places, and it can lead to decreased attention, vigilance and cognitive performance, which is dangerous in the situations such as driving, vessel maneuvering, etc. By directly measuring the neurophysiological activities happening in the brain, electroencephalography (EEG) signal can be used as a good indicator of mental fatigue. A classic EEG-based brain state recognition system requires labeled data from the user to calibrate the classifier each time before the use. For fatigue recognition, we argue that it is not practical to do so since the induction of fatigue state is usually long and weary. It is desired that the system can be calibrated using readily available fatigue data, and be applied to a new user with adequate recognition accuracy. In this paper, we explore performance of cross-subject fatigue recognition algorithms using the recently published EEG dataset labeled with two levels of fatigue. We evaluate three categories of classification method: classic classifier such as logistic regression, transfer learning-enabled classifier using transfer component analysis, and deep-learning based classifier such as EEGNet. Our results show that transfer learning-enabled classifier can outperform the other two for cross-subject fatigue recognition on a consistent basis. Specifically, transfer component analysis (TCA) improves the cross-subject recognition accuracy to 72.70 % that is higher than using just logistic regression (LR) by 9.08 % and EEGNet by 8.72 – 12.86 %.**

Keywords-Cross subject fatigue recognition; EEG; transfer learning; domain adaptation; deep learning

I. Introduction

Resulting from brain over-activity, mental fatigue can happen after long-time continuous concentration on a task, stressful work, or be caused by other factors such as sleep deprivation. It can lead to a decrease in attention, vigilance and cognitive performance, which is dangerous for certain jobs, e.g., pilots, vehicle drivers, helmsmen. For example, the study in [1] shows that fatigue is the main cause of unsafe practices and impaired performances for crane tower operators. Since the onset of mental fatigue bears direct consequences on error chances and safety levels for a wide range of careers, it is desirable to detect mental fatigue in its early stage to prevent such situations in advance.

Fatigue can be measured subjectively by psychometric questionnaires, e.g., the Checklist Individual Strength questionnaire (CIS) [2]. Cognitive tests such as Psycho-motor Vigilance Test (PVT) [3] can also be used to measure fatigue according to the subject's reaction to the visual stimulus. However, such measurements inevitably interrupt the subject's ongoing work, which makes it undesirable when the subject is focusing on the task such as driving a car, or operating a vessel. In addition, such methods also fail to monitor the fatigue state in real-time, which is essential for avoiding accidents in advance. Alternatively, physiological measurements, such as electroencephalography (EEG), electrocardiography (ECG), electrooculography (EOG) or eye tracking systems can be used to detect mental fatigue with a high temporal resolution by nonintrusive ways. Specifically, change of heart rate [4], increase in eye blink rates [5], as well as the increase of the pupil size [6] have all been proved to be valid indicators of mental fatigue. Among these physiological signals, EEG is believed to be the best one for detecting mental fatigue, since it directly measures neurophysiological activities happening in the brain.

Although much existing research has proved the strong relation between EEG pattern and mental fatigue, most

DOI: 10.1109/CW.2019.00048

existing algorithms are based on the intra-subject condition [7-9] – sample EEG signals from the subject under non-fatigue and fatigue state are necessary for training the classifiers. In fact, these methods are not practical for common daily use since mental fatigue is a long and gradual process such that it may take hours to induce fatigue. The challenge of subject-independent fatigue detection lies in the non-stable and non-linear nature of EEG signal – it is different for different subjects and even for the same subject in different trails. In this paper, to evaluate the performance of subject-independent fatigue recognition, we present experiments to test several methods for subject-independent fatigue detection. Three classification methods are compared, namely classic classifier (logistic regression), transfer learning-enabled classifier (transfer component analysis), and deep-learning based classifier (EEGNet). EEG data recorded before demanding events are used in cross-subjects recognition of two fatigue levels.

The paper is organized as follows. Related works are reviewed in Section II. Dataset and methods are described in Section III. The results are presented in Section IV which is followed by the discussion in Section V. The paper is concluded in Section VI.

II. Related works

A. Fatigue detection from EEG signal

Much research has been done to investigate the correlation between fatigue and EEG patterns. It was shown that spectral band power features can be used as a good indicator of different levels of mental fatigue. For example, an overnight sleep-deprived simulated driving task conducted on 12 subjects by Gharagozlou et al. [10] suggested that fatigue could be indicated by increases in α power. Further research showed that fatigue is associated with significant increases in α and $(\theta + \alpha)/\beta$, as well as the decrease in θ/α values [11]. Jap et al. [12] discovered a more general result that the ratio of slow to fast EEG waves increased when the subject experiences fatigue. Chen et al. [13] conducted a 2-hour continuous mental arithmetic task without any break on 12 subjects. Their results showed that mental fatigue leads to increased widespread EEG coherence which is not limited to specific brain regions.

Entropy features of EEG signal have also been found to be valid indicators of mental fatigue. Liu et al. [14] used approximate entropy and Kolmogorov complexity of the EEG signal as discriminators between different fatigue states. In [15], 4 different types of entropy combined with 10 state-of-the-art classifiers were used for subject-specific driver fatigue classification. The best accuracy of 96.60% on a single EEG channel was achieved when using Fuzzy Entropy. In [16], an accuracy of 94.00% for binary fatigue classification on 22 healthy subjects was achieved by using entropy features combined with Gradient Boosting Decision Tree Model.

In addition to traditional methods, recent progress relies on using deep learning models for fatigue classification. In [7], Residual Convolutional Neural Network (EEG-Conv-R) achieved an average accuracy of 84.38% for inter-subject fatigue classification, compared to 81.85 and 75.55% for SVM and LSTM. Data were collected from 10 healthy subjects over 16 channels. In [8], a deep neural network with SVM classifier at the last layer achieved an accuracy of 73.29%.

However, most of the existing works are based on the intra-subject condition – sample EEG signals from the subject under non-fatigue and fatigue state are necessary for training the classifiers. Although several papers [7, 9] declared their inter-subject accuracy, their experiments were designed by mixing up EEG samples from all subjects and dividing them into training and testing sets, which fails to separate EEG data from different individuals for training and testing. That said, in their inter-subject evaluation, part of the test subject's data was used for training. A preliminary work done by [17] achieved an accuracy of 39.80% for 4-level subject-independent fatigue recognition using fractal dimension, 6 statistical features and a linear discriminant analysis classifier without any domain adaptation technique used for the test subject. It is expected that our methods can allow a higher accuracy since the transfer learning techniques allow the classifiers to adapt to the test domain.

B. Cross-subject EEG signal recognition by transfer learning

In a cross-subject EEG recognition task, the training data and test data are from different subjects. Due to the fact that EEG patterns are subject-specific, EEG data and hence the features extracted therefrom tend to distribute dissimilarly among different subjects. Classic machine learning approaches assume that the training data and test data follow the same distribution, which can be hardly satisfied in a cross-subject EEG recognition task. This distribution mismatch often causes degraded recognition accuracy. However, from a practical point of view, cross-subject EEG recognition is desired as it does not require labeled training data collected from the test subject. Considering the mental fatigue recognition task where it takes several hours to induce the fatigue state for the collection of labeled training data, it is not practical to calibrate the classifier each time before we can use it to recognize the fatigue state of a test subject. It is therefore desired that we can train a classifier with a readily available dataset comprising labeled data from other subjects, and apply the trained classifier to a test subject while maintaining the recognition accuracy. Domain adaptation [18, 19], a branch of transfer learning, addresses this concern. In a typical unsupervised domain adaptation problem, we have source domain data $D_S = \{(x_{S_1}, y_{S_1}), \dots, (x_{Sn_1}, y_{Sn_1})\}$ with labels, and target domain data $D_T = \{x_{T_1}, \dots, x_{T_{n_2}}\}$ needing to be classified. Let $\mathcal{P}(X_S)$ and $\mathcal{Q}(X_T)$ be the marginal distributions of X_S and X_T from the source and target domain, respectively. It is assumed that $\mathcal{P} \neq \mathcal{Q}$, but there exists a transformation ϕ such that $P\big(\phi(X_S)\big) \approx P\big(\phi(X_T)\big)$ and $P\big(Y_S \big| \phi(X_S)\big) \approx P\big(Y_T \big| \phi(X_T)\big)$. Using this mapping, classic machine learning methods can be applied, where we can train a classifier on $\phi(X_S)$ with Y_S, and predict the class labels for $\phi(X_T)$. In [20], we showed that domain adaptation can effectively improve the cross-subject recognition accuracy for EEG-based

DOI: 10.1109/CW.2019.00048

emotion recognition. In this paper, we extend the investigation to EEG-based mental fatigue recognition.

III. Materials and Methods

A. Dataset description

In order to test cross-subject fatigue recognition, experiments are conducted on an open dataset. The dataset was published by Cao et al. [21], which were collected during the period from 2005 to 2012. In the experiment, fatigue and drowsiness were induced by a 90-minute sustained-attention night-time driving task in an immersive driving simulator. The participants were asked to drive and maintain the car in the center of the lane. Lane-departure events were randomly induced which makes the car drift to the left or right from the lane, and participants were asked to move back as quickly as possible by steering the wheel. The next event happened in 5–10 seconds after the car returned to the center lane. The participants needed to sustain their attention to the random lane departure events throughout the whole experiment, and their reactions were timed. The reaction time provides a gauge of the subject's fatigue level.

Twenty-seven participants were invited to the experiment. EEG signals were recorded during the whole 90-minute experiment using Quik-Cap (NeuroScan) with 30 valid channels plus 2 reference channels based on a modified international 10–20 system at a sampling rate of 500 Hz. The raw dataset contains 18.21 GB files which are saved in ".set" format. The dataset was released recently and accessible from [22].

B. Data preparation

The preprocessed version of the dataset available from [23] was used in this paper. As described by the authors [21], the raw EEG signals were filtered by 1-Hz high-pass and 50-Hz low-pass finite impulse response (FIR) filters. Apparent eye blinks that contaminate the EEG signals were manually removed through visual inspection by the authors of the dataset. Ocular and muscular artefacts were removed by the Automatic Artifact Removal (AAR) plug-in of EEGLAB [24]. The processed data were finally downsampled to 128 Hz.

As for epoch extraction, we follow the procedures the authors used on the same dataset in their previous paper [25]. Specifically, 3s-long EEG data (epoch) prior to the onset of the lane-departure events were extracted. The fatigue state in this duration was quantitatively estimated based on the reaction time (RT), which was the duration between the onset of the lane-departure event and the onset of the counter-steering event. The RT in each lane-departure event was termed local RT. Additionally, global RT was defined as the average of local RTs across all epochs within a 90-second window before the onset of the deviation event. An alert RT was defined as the 5th percentile of all local RTs in the entire session. Then, the EEG epochs were labeled as such [25]: epochs with both local RT and global RT less than 1.5 × alert RT were non-fatigue epochs; epochs with both local RT and global RT greater than 2.5 × alert RT were fatigue epochs. To ensure sufficient epochs for training and testing, we consider subjects that have at least 50 epochs for both states. Finally, the non-fatigue and fatigue epochs were balanced for each participant. In this way, totally 837 non-fatigue epochs and 837 fatigue epochs from 11 different participants were extracted, and the size of each epoch was 30 (channels) × 384 (sample points). Sixteen subjects were excluded due to not having at least 50 epochs for fatigue/non fatigue states. The epoch numbers for each eligible subject are given in Table I.

Table I Epoch numbers for each eligible subject.

Subject ID	**Number of epochs**	
	Fatigue	***Non fatigue***
1	94	94
5	66	66
22	75	75
31	74	74
35	85	85
41	83	83
42	51	51
43	70	70
44	72	72
45	54	54
53	113	113
Total	837	837

C. Feature extraction

We used the spectral band power as fatigue features in this study, which were widely used in existing EEG-based fatigue-related studies [10-12, 26]. The spectral band power was computed via Fast Fourier Transform on each EEG epoch from these four spectral bands: delta (1 – 4 Hz), theta (4 – 8 Hz), alpha (8 – 12 Hz) and beta (12 – 30Hz). The final feature vector was a concatenation of spectral powers extracted from four bands and all available channels. In this study, the final feature vector was of 4 × 30 = 120 dimensions.

D. Transfer learning method

Transfer Component Analysis (TCA) was proposed by Pan et al. [18] to mitigate the mismatch of distributions between source data and target data, which causes degraded classification accuracy. It seeks a projection to a latent subspace, where the projected source data and target data achieves a reduced Maximum Mean Discrepancy (MMD) in a reproducing kernel Hilbert space (RKHS) [27], which measures the distance between the empirical means of two distributions. It has proven that MMD will asymptotically approach zero if and only if the two distributions are identical when the RKHS is universal [28]. Using the kernel trick, the MMD of source data X_S and target data X_T in the resultant latent subspace evaluates to

$$\mathrm{MMD}(X_S, X_T) = \mathrm{tr}(W^\top KLKW), \quad (1)$$

where $\mathrm{tr}(\cdot)$ is the trace operator, $W \in \mathbb{R}^{(n_1+n_2)\times h}$ is the projection matrix, n_1 and n_2 are the number of examples in X_s and X_t, respectively, h is the dimension of the latent space, $K = [k_{ij}]$ is the kernel matrix defined on $X = [X_S; X_T]$, and $L = [L_{ij}]$ where

DOI: 10.1109/CW.2019.00048

$$L_{ij} = \begin{cases} \frac{1}{n_1^2} \ if\ x_i, x_j \in X_S, \\ \frac{1}{n_2^2} \ if\ x_i, x_j \in X_T, \\ -\frac{1}{n_1 n_2} \ otherwise\,. \end{cases}$$

TCA seeks to minimize the MMD plus a regularization term subjecting to a variance constraint:

$$\min_W \mathrm{tr}(W^\top KLKW) + \mu \mathrm{tr}(W^\top W), \qquad (2)$$
$$\text{s. t. } W^\top KHKW = I,$$

where $H = I - n^{-1}\mathbf{1}_n\mathbf{1}_n^\top$ is a centering matrix, $n = n_1 + n_2$, $\mathbf{1}_n$ is an all-one vector of dimension n, and μ is the trade-off parameter. The solution to W is the h eigenvectors of $(KLK + \mu I)^{-1}KHK$ corresponding to the h leading eigenvalues [18].

E. Baseline methods for comparison

We use logistic regression (LR) as a baseline method. LR is a simple yet effective binary classifier. Given a feature example x, the probability of predicting x as class 1 is estimated by

$$P(y = 1|x;\ W, b) = \frac{1}{\left(1+e^{-(W^\top x+b)}\right)}, \qquad (3)$$

where W and b are the model parameters that need to be fit on the training data, and are usually found by gradient-descent based optimization.

In addition, deep learning models are evaluated for comparison. It is expected that such models have a better performance than traditional machine learning models for cross-subject classification, since they can have a larger capacity with more parameters to accommodate a large amount of data from different subjects. Therefore, a state-of-the-art deep learning model EEGNet specially designed for EEG signal classification is tested for comparison [29]. It was designed for a general EEG signal classification purpose and proved to work for both intra-subject and cross-subject classification on several different BCI-related open datasets including P300 visual-evoked potentials, error-related negativity responses (ERN), movement-related cortical potentials (MRCP), and sensorimotor rhythms (SMR). It is worth mentioning that this model takes raw EEG signals as input and directly learn from raw signals instead of hand-engineered features. The choice of this model was also motivated by the availability of source code published by the authors of [30].

The structure of EEGNet was designed in analogous to the bandpass and CSP spatial filter steps in Filter bank common spatial patterns (FBCSP). EEGNet uses depthwise convolution [31] instead of fully connected layers for the purpose of reducing the number of trainable parameters. It contains a temporal convolution block, a spatial convolution block, a separable convolution block, followed by a dense layer.

F. Classification

To simulate cross-subject fatigue recognition, we carried out leave-one-subject-out cross-validation on the dataset. In a leave-one-subject-out cross-validation setting, the data from one subject were held out from the dataset and reserved as test data, and the data from remaining subjects were pooled together and used as the training data. The recognition accuracy was then evaluated on the held-out subject. The process was repeated until each subject has served as test subject once.

For training deep learning models, we followed the procedures used in the original paper [29] for cross-subject classification. Each time, one subject is selected as the test subject, the other two subjects are used for validation, and the rest are used for training. Test, validation and training subjects are selected in a sequential order. That is, when subject i is tested, subject $i + 1$ and $i + 2$ are used for validation. In this sequence, the last subject is followed by the first subject to form a circle. The process was repeated until each subject has served as test subject once.

G. Setting of hyperparameters

LR: The model parameters (W and b) of LR were optimized by batch gradient descent. The gradient descent was set to run for 100 iterations. The optimization stopped after 100 iterations and the latest W and b were used as model parameters.

TCA: We used a linear kernel K in (2) and set μ to 1. The latent dimension h was set to 80, as it is desired to learn a lower dimensional representation than the original features (120 dimensional).

EEGNet: We use the default parameter setting implemented in [30]. Both of the EEGNet models – EEGNet-4,2 and EEGNet-8,2, as proposed in the original paper [29] were tested. The models were fit by minimizing the binary cross-entropy using Adam optimizer with default parameters. We trained the models for 100 epochs with a minibatch size of 100. The dropout rate was 0.5 for both models. The number of trainable parameters for EEGNet-4,2 and EEGNet-8, 2 are 946 and 1954, respectively. The structure and parameters of both networks are listed in Table II.

IV. RESULTS

The mean accuracy of all subjects in the leave-one-subject-out cross-validation experiment are presented in Table III. TCA + LR gives the best accuracy of 72.70 %. Notably, applying TCA before LR can effectively improve the accuracy from 63.62 % to 72.70 % (9.08 % difference). TCA + LR also gives the smallest standard deviation, signifying that the improvement is fairly consistent across different test subjects. This can be seen in Fig. 1 where the recognition accuracy for each individual subject is shown. TCA + LR consistently gives better performance than LR on all subjects except subject 7. The two deep learning-based methods do not perform better than classic machine learning approach in this experiment. EEGNet-8, 2 performs similarly as LR, whereas EEGNet-4, 2 performs slightly worse.

DOI: 10.1109/CW.2019.00048

TABLE II EEGNET STRUCTURE AND NUMBER OF PARAMETERS.

Layer	EEGNet - 8, 2		EEGNet - 4, 2	
	Output shape	***Params***	***Output shape***	***Params***
Input	(30, 384, 1)	0	(30, 384, 1)	0
Conv2D	(30, 384, 8)	512	(30, 384, 4)	256
Batch Normalization	(30, 384, 8)	120	(30, 384, 4)	120
Depthwise Conv2D	(1, 384, 16)	480	(1, 384, 8)	240
Batch Normalization	(1, 384, 16)	4	(1, 384, 8)	4
Activation	(1, 384, 16)	0	(1, 384, 8)	0
Average Pooling 2D	(1, 96, 16)	0	(1, 96, 8)	0
Dropout	(1, 96, 16)	0	(1, 96, 8)	0
Separable Con2D	(1, 96, 16)	512	(1, 96, 8)	192
Batch Normalization	(1, 96, 16)	4	(1, 96, 8)	4
Activation	(1, 96, 16)	0	(1, 96, 8)	0
Average Pooling 2D	(1, 12, 16)	0	(1, 12, 8)	0
Dropout	(1, 12, 16)	0	(1, 12, 8)	0
Flatten	192	0	96	0
Dense	2	386	2	194
Activation	2	0	2	0
Total Params		2018		1010
Trainable Params		1954		946
Non-trainable Params		64		64

We applied one-way ANOVA to analyze the significance of difference of mean accuracy. The analysis showed a significant difference in the mean accuracy of different methods ($F_{(3,40)}$=3.25, p=0.03). Post-hoc Bonferroni-corrected pairwise comparison suggested that there was a significant difference between TCA + LR and EEGNet-4, 2 (p=0.027). Other pairs of method did not exhibit significant difference (p>0.05).

V. DISCUSSION

Contrary to our expectation, the deep learning-based method did not give better recognition accuracy than classical machine learning approaches for fatigue recognition in this dataset. The reason might be due to directly learning from raw EEG signals. Raw EEG signals tend to be noisier than the features extracted therefrom. Without the infusion of expert knowledge (in the form of hand-engineered features) into the deep neural network, it might be learning more from the noisy components of EEG than the informative components. Another reason might be due to the relatively small amount of training samples. Unlike deep learning for image recognition which learns from abundant training samples, EEG recognition is often plagued by a small-sized training set. The deep learning models have a greater capacity due to the large number of parameters. Without sufficiently large training samples, the deep neural network might fail to learn meaningful features.

TABLE III MEAN ACCURACY (%) AND STANDARD DEVIATION (%) OF LEAVE-OUT-SUBJECT-OUT CROSS-VALIDATION FOR RECOGNIZING FATIGUE AND NONFATIGUE STATES

Method	Mean	Standard deviation
LR	63.62	9.93
TCA + LR	**72.70**	**9.42**
EEGNet-8,2	63.98	11.10
EEGNet-4,2	59.84	9.55

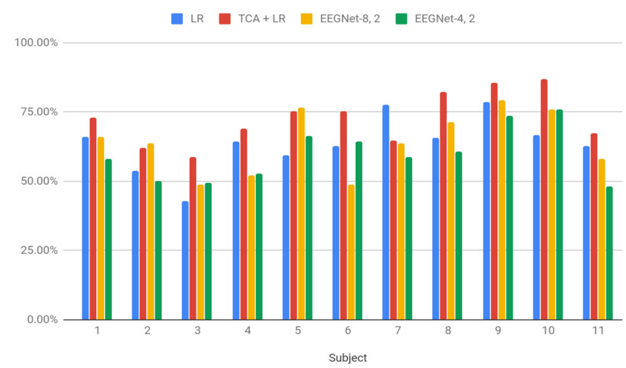

Figure 1 Recognition accuracy for each individual subject

VI. CONCLUSION

In this paper, we present a study of EEG-based cross-subject mental fatigue recognition. The typical EEG signal recognition method requires that the subject provide labeled training data to calibrate the classifier each time before use. However, for fatigue recognition, we argue that it is not practical to induce fatigue states and record the respective EEG signals for calibration for every subject, considering that the fatigue induction is long and weary. Cross-subject fatigue recognition could be the solution to a "plug and play" EEG-based fatigue recognition system. We examined and compared several methods against each other: classic classifier (LR), transfer learning-enabled classifier (TCA + LR), and deep learning-based method (EEGNet). Experiments were done on a public EEG dataset where each subject was instructed to drive a car in a simulator for 90 minutes. Lane departure events randomly occurred during the course of driving. The fatigue state was gauged by the time of reaction to the lane departure event. Leave-out-subject-out cross-validation was adopted to simulate the cross-subject fatigue recognition task. We found that TCA + LR performed the best among all methods: significantly better than EEGNet-4, 2 and presumably better than LR. Thus, the transfer learning-enabled classifier obtained the best cross-subjects accuracy, which outperformed logistic regression (LR) by 9.08 % and EEGNet by 8.72 – 12.86 %. This suggests that transfer learning-enabled classifier may be a promising method for cross-subject fatigue recognition. We observed that deep learning-based method does not perform better than classic

DOI: 10.1109/CW.2019.00048

approaches, possibly due to the limited training samples and the noisy EEG components.

ACKNOWLEDGMENT

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative.

REFERENCES

[1] V. W. Y. Tam and I. W. H. Fung, "Tower crane safety in the construction industry: A Hong Kong study," *Safety Science,* vol. 49, no. 2, pp. 208-215, 2011.

[2] A. J. H. M. Beurskens, U. Bültmann, I. Kant, J. H. M. M. Vercoulen, G. Bleijenberg, and G. M. H. Swaen, "Fatigue among working people: validity of a questionnaire measure," *Occupational and Environmental Medicine,* vol. 57, no. 5, p. 353, 2000.

[3] M. Basner and J. Rubinstein, "Fitness for duty: a 3-minute version of the Psychomotor Vigilance Test predicts fatigue-related declines in luggage-screening performance," *Journal of occupational and environmental medicine,* vol. 53, no. 10, pp. 1146-1154, 2011.

[4] Y. Tran, N. Wijesuriya, M. Tarvainen, P. Karjalainen, and A. Craig, "The relationship between spectral changes in heart rate variability and fatigue," *Journal of Psychophysiology,* vol. 23, no. 3, pp. 143-151, 2009.

[5] S. Hu and G. Zheng, "Driver drowsiness detection with eyelid related parameters by Support Vector Machine," *Expert Systems with Applications,* vol. 36, no. 4, pp. 7651-7658, 2009.

[6] Y. Morad, H. Lemberg, N. Yofe, and Y. Dagan, "Pupillography as an objective indicator of fatigue," *Current eye research,* vol. 21, no. 1, pp. 535-542, 2000.

[7] H. Zeng, C. Yang, G. Dai, F. Qin, J. Zhang, and W. Kong, "EEG classification of driver mental states by deep learning," *Cognitive neurodynamics,* vol. 12, no. 6, pp. 597-606, 2018.

[8] P. P. San, S. H. Ling, R. Chai, Y. Tran, A. Craig, and H. Nguyen, "EEG-based driver fatigue detection using hybrid deep generic model," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 800-803.

[9] R. Chai *et al.*, "Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system," *IEEE journal of biomedical and health informatics,* vol. 21, no. 3, pp. 715-724, 2016.

[10] F. Gharagozlou *et al.*, "Detecting driver mental fatigue based on EEG alpha power changes during simulated driving," *Iranian journal of public health,* vol. 44, no. 12, p. 1693, 2015.

[11] T. Cao, F. Wan, C. M. Wong, J. N. da Cruz, and Y. Hu, "Objective evaluation of fatigue by EEG spectral analysis in steady-state visual evoked potential-based brain-computer interfaces," *BioMedical Engineering OnLine,* vol. 13, no. 1, p. 28, March 12 2014.

[12] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Systems with Applications,* vol. 36, no. 2, Part 1, pp. 2352-2359, 2009.

[13] L. Chen, Y. Zhao, J. Zhang, and J. Zou, "Investigation of Mental Fatigue Induced by a Continuous Mental Arithmetic Task Based on EEG Coherence Analysis," in *Life System Modeling and Simulation*, 2014, pp. 33-38.

[14] J. Liu, C. Zhang, and C. Zheng, "EEG-based estimation of mental fatigue by using KPCA–HMM and complexity parameters," *Biomedical Signal Processing and Control,* vol. 5, no. 2, pp. 124-130, 2010.

[15] J. Hu, "Comparison of different features and classifiers for driver fatigue detection based on a single EEG channel," *Computational and mathematical methods in medicine,* vol. 2017, 2017.

[16] J. Hu and J. Min, "Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model," *Cognitive neurodynamics,* vol. 12, no. 4, pp. 431-440, 2018.

[17] Y. Liu, Z. Lan, H. H. G. Khoo, K. H. H. Li, O. Sourina, and W. Mueller-Wittig, "EEG-based Evaluation of Mental Fatigue Using Machine Learning Algorithms," in *International Conference on Cyberworlds (CW)*, 2018, pp. 276-279.

[18] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks,* vol. 22, no. 2, pp. 199-210, 2011.

[19] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering,* vol. 22, no. 10, pp. 1345-1359, 2010.

[20] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain Adaptation Techniques for EEG-Based Emotion Recognition: A Comparative Study on Two Public Datasets," *IEEE Transactions on Cognitive and Developmental Systems,* vol. 11, no. 1, pp. 85-94, 2019.

[21] Z. Cao, C.-H. Chuang, J.-K. King, and C.-T. Lin, "Multi-channel EEG recordings during a sustained-attention driving task," *Scientific data,* vol. 6, 2019.

[22] Z. Cao, M. Chuang, J. T. King, and C.-T. Lin. (2019). *Multi-channel EEG recordings during a sustained-attention driving task (raw dataset)*. Available: https://figshare.com/articles/Multi-channel_EEG_recordings_during_a_sustained-attention_driving_task/6427334/5

[23] Z. Cao, M. Chuang, J. T. King, and C.-T. Lin. (2019). *Multi-channel EEG recordings during a sustained-attention driving task (pre-processed dataset)*. Available: https://figshare.com/articles/Multi-channel_EEG_recordings_during_a_sustained-attention_driving_task_preprocessed_dataset_/7666055

[24] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *Journal of neuroscience methods,* vol. 134, no. 1, pp. 9-21, 2004.

[25] C.-S. Wei, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "Toward drowsiness detection using non-hair-bearing EEG-based brain-computer interfaces," *IEEE transactions on neural systems and rehabilitation engineering,* vol. 26, no. 2, pp. 400-406, 2018.

[26] L. J. Trejo, K. Kubitz, R. Rosipal, R. L. Kochavi, and L. D. Montgomery, "EEG-Based Estimation and Classification of Mental Fatigue," *Psychology,* vol. 6, no. 5, p. 18, 2015.

[27] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," in *Advances in neural information processing systems*, 2007, pp. 513-520.

[28] A. Smola, A. Gretton, L. Song, and B. Schölkopf, "A Hilbert space embedding for distributions," in *International Conference on Algorithmic Learning Theory*, 2007, pp. 13-31.

[29] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of neural engineering,* vol. 15, no. 5, p. 056013, 2018.

[30] V. J. Lawhern. (2019). *ARL-EEGmodels*. Available: https://github.com/vlawhern/arl-eegmodels

[31] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251-1258.

DOI: 10.1109/CW.2019.00048