Preprint Version. Manuscript submitted to and accepted by 2018 International Conference on Cyberworlds.

EEG-based Evaluation of Mental Fatigue Using Machine Learning Algorithms

Yisi Liu, Zirui Lan
Fraunhofer Singapore
Singapore
{liuys, lanz0001}@ntu.edu.sg

Han Hua Glenn Khoo, King Ho Holden Li
Nanyang Technological University
Singapore
{gkhoo003, holdenli}@ ntu.edu.sg

Olga Sourina, Wolfgang Mueller-Wittig
Fraunhofer Singapore
Singapore
eosourina@ntu.edu.sg,
wolfgang.mueller-wittig@fraunhofer.sg

***Abstract*—When people are exhausted both physically and mentally from overexertion, they experience fatigue. Fatigue can lead to a decrease in motivation and vigilance which may result in certain accidents or injuries. It is crucial to monitor fatigue in workplace for safety reasons and well-being of the workers. In this paper, Electroencephalogram (EEG)-based evaluation of mental fatigue is investigated using the state-of-the-art machine learning algorithms. An experiment lasted around 2 hours and 30 minutes was designed and carried out to induce four levels of fatigue and collect EEG data from seven subjects. The results show that for subject-dependent 4-level fatigue recognition, the best average accuracy of 93.45% was achieved by using 6 statistical features with a linear SVM classifier. With subject-independent approach, the best average accuracy of 39.80% for 4 levels was achieved by using fractal dimension, 6 statistical features and a linear discriminant analysis classifier. The EEG-based fatigue recognition has the potential to be used in workplace such as cranes to monitor the fatigue of operators who are often subjected to long working hours with heavy workloads.**

Keywords- EEG; machine learning; fatigue recognition

I. INTRODUCTION

Fatigue is common in daily life and it could be a hazard in workplace which may result in accidents and injuries due to low vigilance and efficiency. For example, in the construction industry, the workers are often subjected to long working hours with heavy workloads. It is claimed in [1, 2] that fatigue could be one of the reasons that causes human errors. It is also shown in [3] that through seven-hour driving, the driver experienced fatigue and gradually the performance was decreased. Thus, it is important to evaluate/monitor fatigue level of the workers in workplace or the drivers who must continuously drive for a relatively long time.

Usually questionnaires [4, 5], cognitive tests [6, 7], or bio-signals [8] are used to evaluate fatigue. Compared to questionnaires and cognitive tests which interrupt the ongoing work or only are given at the end of tasks, bio-signals based fatigue recognition can provide continuous monitoring and does not have to interfere with the subjects' primary tasks.

In this paper, we design and implement an experiment to induce 4 levels of fatigue and propose both subject-dependent and subject-independent algorithms using the state-of-the-art machine learning techniques.

The paper is constructed as follows. In Section II, the definition of fatigue, experiment of fatigue evocation, and the state-of-the-art EEG-based fatigue recognition algorithms are reviewed. In Section III, the experiment design is introduced. In Section IV, the results of proposed subject-dependent and subject-independent data analyses are presented. Section V concludes the paper.

II. RELATED WORK

A. Fatigue

Fatigue refers to an extreme way of tiredness that could result from overexerting oneself mentally, physically or both [1]. It could cause a loss of efficiency with a lack of effort [2]. It may also lead to a decrease in motivation and vigilance in a being, which results in a higher probability of accidents or injuries.

Fatigue is usually measured by survey questionnaires [3, 4]. The subjects are given a series of questions regarding how they feel. For example, [5] validates the use of the Checklist Individual Strength questionnaire (CIS) to measure fatigue among working people. A checklist is included in this questionnaire and the subjects need to rate the degree to which they agree to the questions such as "I feel tired", "I feel very active". Besides questionnaires, a cognitive test such as Psycho-motor Vigilance Test (PVT) is used in research to evaluate fatigue level. In the PVT test, the subjects react to the visual stimuli and the reaction time is recorded as a measurement of fatigue [6, 7].

Besides questionnaires and cognitive tests, in recent studies, several other methods are also used as the measurements of fatigue such as Electroencephalogram (EEG), Skin Conductance Response (SCR), heart rate monitoring, and oxygen intake [8]. The EEG-based recognition outperforms the others as it has high temporal resolution. Additionally, in [9], it is found that emotions also correlate to fatigue, and [10] shows that the EEG-based brain state recognition can identify states such as emotion and workload with good accuracy. Thus, in this paper EEG-based fatigue recognition is studied.

B. Review on Experiments

There are different ways to induce fatigue under the experimental settings, namely, the trail making test, mirror drawing test and n-back tests. Such experiments usually last for a duration of 2 hours continuously to guarantee the elicitation of fatigue [11]. The trail making test consists of circles numbering, for example, from 1-25, and the numbers located at random positions on the screen [12]. The subjects need to click the numbers in sequence as fast as possible. To get fatigue, the subjects have to do the test at least for 30 minutes. The mirror drawing test requires subjects to draw and to trace pictures by looking at the inverted image of the pictures through a small glass mirror [13]. To get fatigue, a certain period for at least 30 minutes is required. Lastly, the n-back test requires subjects to use their short-term memory. They need to memorize the letters flashed on the screen and indicate by pressing a certain key when the current letter matches the one from n steps earlier in the sequence [14, 15]. The difficulty level of the n-back test can be changed by increasing the number of "n" defined in the test such as 1-back test and 2-back test. It can also include two senses concurrently, such as visual and audio senses. For example, besides letters flash on the screen, the reading of alphabets is given at the same time, and the subjects need to react to both stimuli. In this paper, the n-back test is selected since it could be used to invoke different levels of fatigue by various difficulty levels.

C. EEG-base Fatigue Recognition

The correlation between fatigue and EEG has been studied in the literatures. Jap et al. [16] recruited 52 subjects and recorded their EEG signals during a monotonous driving session. They examined the spectral components of EEG and established that the ratio of slow waves (e.g., α, θ waves) to fast waves (e.g., β waves) would increase when the subject was experiencing fatigue. Chen et al. [17] carried out an experiment on 12 subjects to induce fatigue by a 2-hour mental arithmetic task without any break. They investigated the spectral coherence among 28 electrode pairs within four frequency bands (delta, theta, alpha, and beta), and proved that mental fatigue was accompanied by increased EEG coherence. Cao et al. [18] investigated the fatigue caused by using a steady state visually evoked potential (SSVEP)-based brain-computer interface, and reported consistent observations of significant increases in α and $(\alpha+\theta)/\beta$ ratio, as well as the decreases in θ/α ratio when the subjects were suffering fatigue. Gharagozlou et al. [19] proposed to induce fatigue by an overnight sleep-deprived stimulated driving task on 12 subjects, and demonstrated that a significant increase in the absolute α power indicated the onset of fatigue. Liu et al. [20] showed that approximate entropy and Kolmogorov complexity of the EEG signal are indicative of different fatigue states, and proposed to classify the fatigue state with a combination of kernel principal component analysis and Hidden Markov Model using the said features. They reported the best accuracy 84.00% for differentiating fatigue and nonfatigue states. Trejo et al. [21] justified the increase of α and θ spectral power when fatigue kicks in, and classified the fatigue-nonfatigue states with a kernel partial least squares regression classifier, reporting the mean recognition accuracy of 98.80%. Mu et al. [22] used fuzzy entropy as features and support vector machine as classifier, and attain the accuracy 85.00% for binary fatigue level classification.

In this paper, we propose methods that are different from the existing studies [20-22] in two folds. Firstly, we attempt to recognize up to four fatigue levels comparing to the existing works which mainly focus on binary fatigue level differentiation. Secondly, we have taken both subject-dependent and subject-independent approaches while the existing studies mainly evaluate the classification performance on a subject-dependent basis.

III. Experiment Design

An experiment was designed and carried out to elicit fatigue.

A. Subjects

A total of seven males with an average age of 24 have participated in this experiment. None of them has any history of mental illness. All the subjects have been told not to have any caffeine or energy drinks prior to the experiment. In addition, the experiment has been always conducted after lunch, in a dimly lit room with minimal external noise.

B. Procedure

The entire experiment lasted around 2 hours and 30 minutes. After the subjects reached the lab, they were briefly explained about the experiment, and the EEG device was mounted on their head. Then, the two-hour task was initiated, and it included 30 minutes of relaxation (Phase 1), followed by a 1.5-hour continuous n-back test of difficult level 1, 2 and 3 (Phase 2, 3 and 4) with a 10-minute break after the 1-hour mark. In phase 1, the subjects were required to listen to soothing, relaxing music without any sort of distraction, which may include noise or even excessive light. In phase 2-4, the n-back test was employed where a sequence of letters were given to the subjects. The difficulty level for the tests increases with every 30-minute session, namely the 1-back test, 2-back test and the dual 2-back test. For dual 2-back test, the subjects needed to react to not only visual stimuli (letters flashed on the screen) but also audio stimuli (reading of alphabets) in the 2-back manner.

The EEG during each phase was recorded. After every phase, a CIS questionnaire was given to the subjects for them to rate their level of fatigue. It contains 20 questions regarding the subjective feeling of fatigue, concentration, motivation, and physical activity. A final score can be calculated according to the answers to the questionnaire. The higher the points from the questionnaire, the more fatigue the subject experienced.

C. EEG Device

A 14-channel Emotiv device [23] was used to record the EEG data during the experiment. The channels of Emotiv are AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4, locating at frontal lobe, temporal lobe, parietal lobe, and occipital lobe.

IV. Results and Analyses

A. *CIS Questionniare*

From the results of CIS questionnaire shown in Table I, the fatigue level increased with the difficulty level of n-back test deepened.

TABLE I. VAS Questionniare Results

Subject #	Exp. Phase	CIS Score	Subject #	Exp. Phase	CIS Score
1	1	37	4	3	88
	2	56		4	91
	3	106	5	1	37
	4	126		2	56
2	1	67		3	70
	2	86		4	99
	3	91	6	1	60
	4	109		2	85
3	1	71		3	94
	2	81		4	108
	3	83	7	1	70
	4	89		2	78
4	1	67		3	81
	2	78		4	85

B. *EEG-based Fatigue Recognition*

Then the CIS score was used as ground truth to label the EEG data to validate the fatigue recognition algorithm. As a result, EEG data labeled with 4 different levels of fatigue were obtained. Based on the labeled EEG data, we evaluated the recognition accuracy for 4 fatigue levels. We adopted and compared several state-of-the-art EEG features (Statistics [24-26], higher order crossing [25, 27], fractal dimension [24, 25], hjorth [28, 29], signal energy [30], and spectral power [31-33]) and different classifiers (logistic regression, linear discriminant analysis, 1-nearest neighbor, linear support vector machine, and naïve Bayes), and carried out both subject-dependent recognition and subject-independent recognition on a per-subject basis. For subject-dependent recognition, we applied five-fold cross-validation to each subject as was used in [20, 21], where four folds were used as training data and the remaining fold as test data. Four-second window was used for the features extraction. The recognition accuracy averaged across all subjects is presented in Table II. The best average recognition accuracy, 93.45%, was attained by using 6 statistical features with a linear SVM classifier. Naïve Bayes classifier yielded below-par recognition accuracy irrespective of the feature used. SE gave the worst performance except when used with 1-NN classifier. Comparing to existing studies [20-22], our methods can recognize four levels of fatigue while maintaining comparable accuracy to binary fatigue level differentiation. Though subject-dependent fatigue recognition could potentially achieve satisfactory accuracy, the performance was at the cost of prolonged training time – 2 hours' fatigue induction. In real applications, it is not practicable to spend so much time calibrating the algorithm. Subject-independent fatigue recognition could be adopted to eliminate the wearying fatigue induction for the subject-of-interest, by using the labeled EEG data of other subjects to train the classifier. We stress that such research has been lacking in existing studies, and it is crucial to the real application. For subject-independent recognition, we adopted the leave-one-subject-out cross-validation scheme, where we reserved the data from one subject as the test data and pooled together the data from all other subjects as training data, until each subject has served as the test subject once. The recognition accuracy averaged across all subjects presented in Table III. A drastic reduction in recognition accuracy was observed when compared to subject-dependent fatigue recognition presented in Table II. The best accuracy, 39.80%, was attained by fractal dimension + 6 statistical features and a linear discriminant analysis classifier. The same feature also performed the best when used with a logistic regression classifier. Linear SVM, which was found to be the best-performing classifier in subject-dependent application, yielded below-par performance in subject-independent case. One-nearest neighbor and naïve Bayes produced similar recognition results, performing at mediocre level. It is worth highlighting that the subject-independent fatigue recognition is more applicable in real application scenario as it does not require the prolonged, tedious fatigue evocation before the subject can be monitored by such system. Considering its application values, in the future, we will focus on how to further improve the accuracy of subject-independent fatigue recognition.

TABLE II. Subject-dependent recognition accuracy (%)

Feature	Classifier				
	LR	***LDA***	***1-NN***	***LinSVM***	***NB***
6 STAT	**88.80**	88.80	74.04	**93.45**	47.33
36 HOC	72.02	70.11	41.65	66.31	50.89
FD+6STAT+36HOC	81.61	88.02	48.59	88.42	**54.58**
FD+6STAT	88.60	**89.01**	72.49	93.24	47.72
HJORTH	74.51	73.74	69.79	67.93	37.79
SE	58.61	43.48	83.87	37.75	31.08
POWER	67.37	56.51	**84.59**	56.19	33.38

TABLE III. Subject-independent recognition accuracy (%)

Feature	Classifier				
	LR	***LDA***	***1-NN***	***LinSVM***	***NB***
6 STAT	38.59	39.59	31.67	28.01	26.74
36 HOC	33.73	33.97	28.74	26.73	**34.37**
FD+6STAT+36HOC	35.37	38.91	31.40	**29.83**	32.25
FD+6STAT	**39.07**	**39.80**	31.93	25.23	26.65
HJORTH	32.80	33.25	32.35	23.39	24.72
SE	27.73	27.56	27.87	25.09	24.66
POWER	29.00	29.62	**35.64**	25.76	25.17

V. Conclusion

In this paper, we designed and carried out an experiment to induce four levels of fatigue on seven subjects and collected their EEG data while the fatigue induction was ongoing. We proposed methods to recognize four levels of fatigue, with the average accuracy of 93.45% achieved on a subject-dependent basis. The results are more promising in comparison to other works which reported accuracy for 2 classes fatigue state differentiation only. We also took the subject-independent approach to recognizing the 4 fatigue levels, yielding the average accuracy of 39.80%. While the subject-dependent approach could achieve much higher recognition accuracy, such benefit came at the cost of prolonged training time (2 hours) which is not practicable in real application scenarios. The subject-independent approach makes more practical sense but has been lacking sufficient research thus far and it needs more data collected from different subjects to improve accuracy. In the future, we will focus on how to further improve the accuracy of subject-independent fatigue recognition.

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] G. Lewis and S. Wessely, "The epidemiology of fatigue: more questions than answers," Journal of Epidemiology and Community Health, vol. 46, no. 2, pp. 92-97, 1992.

[2] K. Sadeghniiat-Haghighi and Z. Yazdi, "Fatigue management in the workplace," Industrial psychiatry journal, vol. 24, no. 1, pp. 12-17, 2015.

[3] K. A. Lee, G. Hicks, and G. Nino-Murcia, "Validity and reliability of a scale to assess fatigue," Psychiatry research, vol. 36, no. 3, pp. 291-298, 1991.

[4] T. Chalder et al., "Development of a fatigue scale," Journal of psychosomatic research, vol. 37, no. 2, pp. 147-153, 1993.

[5] A. J. Beurskens, U. Bültmann, I. Kant, J. H. Vercoulen, G. Bleijenberg, and G. M. Swaen, "Fatigue among working people: validity of a questionnaire measure," Occupational and environmental medicine, vol. 57, no. 5, pp. 353-357, 2000.

[6] 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, p. 1146, 2011.

[7] Y. Y. Tan, S. T. Lin, and F. Tey, "Development of Fatigue-Associated Measuresment to Determine Fitness for Duty and Monitor Driving Performance," in International Conference on Augmented Cognition, 2013, pp. 608-617.

[8] A. Aryal, A. Ghahramani, and B. Becerik-Gerber, "Monitoring fatigue in construction workers using physiological measurements," Automation in Construction, vol. 82, pp. 154-165, 2017.

[9] G. Lewis and S. Wessely, "The epidemiology of fatigue: more questions than answers," Journal of epidemiology and community health, vol. 46, no. 2, p. 92, 1992.

[10] X. Hou, Y. Liu, O. Sourina, and W. Mueller-Wittig, "CogniMeter: EEG-based emotion, mental workload and stress visual monitoring," in Cyberworlds International Conference on, 2015, pp. 153-160.

[11] 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 International Conference on Life System Modeling and Simulation and International Conference on Intelligent Computing for Sustainable Energy and Environment, 2014, pp. 33-38.

[12] K. Mizuno and Y. Watanabe, "Utility of an advanced trail making test as a neuropsychological tool for an objective evaluation of work efficiency during mental fatigue," in Fatigue science for human health: Springer, 2008, pp. 47-54.

[13] K. Yoshiuchi et al., "Hemodynamic and endocrine responsiveness to mental arithmetic task and mirror drawing test in patients with essential hypertension," American journal of hypertension, vol. 10, no. 3, pp. 243-249, 1997.

[14] T. S. Braver, J. D. Cohen, L. E. Nystrom, J. Jonides, E. E. Smith, and D. C. Noll, "A Parametric Study of Prefrontal Cortex Involvement in Human Working Memory," NeuroImage, vol. 5, no. 1, pp. 49-62, 1997.

[15] M. Tanaka et al., "Impaired selective attention caused by mental fatigue," J Neurol Sci (Turkish), vol. 29, no. 3, pp. 542-553, 2012.

[16] 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.

[17] 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.

[18] 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, 2014.

[19] 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, pp. 1693-1700, 2015.

[20] 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.

[21] 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, pp. 572-589, 2015.

[22] Z. Mu, J. Hu, and J. Yin, "Driving Fatigue Detecting Based on EEG Signals of Forehead Area," International Journal of Pattern Recognition and Artificial Intelligence, vol. 31, no. 5, 2017.

[23] Emotiv. http://www.emotiv.com.

[24] Y. Liu and O. Sourina, "EEG Databases for Emotion Recognition," in Proc. 2013 Int. Conf. on Cyberworlds, pp. 302-309, 2013.

[25] Y. Liu and O. Sourina, "Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm," in Transactions on Computational Science XXIII: Springer, 2014, pp. 199-223.

[26] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE transactions on pattern analysis and machine intelligence, vol. 23, no. 10, pp. 1175-1191, 2001.

[27] P. C. Petrantonakis and L. J. Hadjileontiadis, "Emotion recognition from EEG using higher order crossings," IEEE Transactions on Information Technology in Biomedicine, vol. 14, no. 2, pp. 186-197, 2010.

[28] B. Hjorth, "EEG analysis based on time domain properties," Electroencephalography and clinical neurophysiology, vol. 29, no. 3, pp. 306-310, 1970.

[29] K. Ansari-Asl, G. Chanel, and T. Pun, "A channel selection method for EEG classification in emotion assessment based on synchronization likelihood," in 15th European Signal Processing Conference, 2007, pp. 1241-1245.

[30] F. Feradov and T. Ganchev, "Ranking of EEG time-domain features on the negative emotions recognition task," Annual Journal of Electronics, vol. 9, pp. 26-29, 2015.

[31] K. Ishino and M. Hagiwara, "A feeling estimation system using a simple electroencephalograph," in IEEE International Conference on Systems, Man and Cybernetics, 2003, vol. 5, pp. 4204-4209.

[32] K. Schaaff and T. Schultz, "Towards an EEG-based emotion recognizer for humanoid robots," in IEEE International Workshop on Robot and Human Interactive Communication, 2009, pp. 792-796.

[33] Y. Liu and O. Sourina, "EEG-based Dominance Level Recognition for Emotion-enabled Interaction," in The IEEE International Conference on Multimedia & Expo, 2012, pp. 1039-1044.