

“© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Combining ICA Clustering and Power Spectral Density for Feature Extraction of Mental Fatigue of Spinal Cord Injury Patients

Rifai Chai, *Senior Member, IEEE*, Yvonne Tran, Sai Ho Ling, *Senior Member, IEEE*, Ashley Craig and Hung T. Nguyen, *Senior Member, IEEE*

Abstract— This paper presents the combination of ICA clustering using ICASSO and power spectral density (PSD) as a features extractor of mental fatigue from spinal cord injury (SCI) patients. Initially, the comparison healthy vs. SCI is tested on alert and fatigue data shows that they are comparable. Further, the coefficient determination (R^2) is calculated for testing the variation of data alert vs. fatigue on the SCI group and resulted in a lower R^2 for proposed combination of ICASSO and PSD method compared to the PSD method only. With the lower R^2 values, this shows that the proposed method ICASSO and PSD is able to give more distinction for separating fatigue vs. alert data variation. The statistical significance is found across four EEG bands and EEG channels.

I. INTRODUCTION

Spinal cord injury (SCI) is a damage to the spinal cord that can be caused by a disease or traumatic injury which contributes to a bruised or severing of the spinal cord. The further impact of SCI can lead to reduced mobility, psychological aspect, cognitive function, pain and fatigue which affects the quality of life for a patient with SCI. Mental fatigue is a common issue in SCI which related to the mental state of tiredness with a decrease of alertness and motivation. The effect of the mental fatigue leads to feeling of unwillingness to continue to perform a task and negatively affects the ability of the SCI patient to plan and prepare to execute a task such as using assistive technology and rehabilitation [1]. Therefore, it is crucial to be able to distinguish between the mental fatigue state and non-fatigue state (alert) in order to provide an optimum outcome when the SCI person uses the assistive technology and rehabilitation system.

In general, there are two main strategies which can be used for assessing mental fatigue including psychological and physiological strategies. The psychological strategy is popular in mental health research, based on individual responses to the psychometric questionnaire which can be

biased to a particular individual as it contains subjective responses. As a result, this method may not be a practical solution for monitoring mental fatigue. The other strategy is based on physiological measurement including: (1) video monitoring to capture the facial and eye tracking for fatigue occurrence, (2) heart signal measurement using electrocardiogram, (3) eyes movement/activity measurement using electrooculogram (EOG), and (4) brain signal measurement using electroencephalogram (EEG). This paper concentrates on exploring the mental fatigue analysis based on the measurement of the brain signal using the EEG. The advantages of the EEG measurement are that it can provide a direct measurement of neurophysiological activity related to the mental fatigue occurrence. In addition, the EEG system is based on non-invasive technology, is portable, cost effective and can be designed for an automatic counter measurement system [2, 3].

Overall, for analysing the EEG signal, the feature extraction module performs a crucial task in order to convert the raw EEG data into a representative feature. The popular strategy for extracting the EEG features is by employing the power spectral density (PSD) algorithm which is able to convert the time domain EEG data into a frequency representation. Then, the PSD data is grouped into four EEG bands, which are delta (0-3Hz), theta (4-7Hz), alpha (8-13Hz) and beta bands (14-30Hz). As novelty, this paper combines the clustering-based independent component analysis (ICA), called ICASSO [4] and the PSD method for extracting the mental fatigue features of SCI patients' data. The clustering-based ICA is able to tackle the ICA by performing the algorithm several times with different conditions and visualising the clustering features of the algorithm to provide reliable components.

II. METHODOLOGY

A. General Structure

The general structure for analyzing the mental fatigue of the EEG data recorded is shown in Fig.1. The first module is data collected from the experiments of mental fatigue of 10 ten SCI patients by using 26 channels EEG. This is continued with the second module for extracting the features from raw 26 EEG channels. From the experiment, the data is divided into two group, fatigue and alert data. The clustering-based ICA using ICASSO is applied to the EEG data (fatigue and alert data) which resulted in ICASSO components. The power spectral density (PSD) is calculated from each EEG channels of the data. The total power of the

Rifai Chai and Hung T. Nguyen is with Faculty of Science, Engineering and Technology, Swinburne University of Technology, Hawthorn, VIC 3122, Australia (e-mail: rchai@swin.edu.au; hungnguyen@swin.edu.au).

Yvonne Tran is Australian Institute of Health Innovation, Macquarie University, Macquarie Park NSW 2109 (e-mail: yvonne.tran@mq.edu.au).

Sai Ho Ling, is with School of Biomedical Engineering, Faculty of Engineering and Information Technology, University of Technology, Sydney, Broadway NSW 2007, Australia (e-mail: Steve.Ling@uts.edu.au).

Ashley Craig is with the Kolling Institute of Medical Research, Sydney Medical School, The University of Sydney (e-mail: a.craig@sydney.edu.au).

PSD is applied to four EEG bands (delta: 0-3Hz; theta: 4-7Hz; alpha: 8-13Hz and beta: 14-30Hz). On the third module, statistical analysis using coefficient determination, R -squared and p -value are used for comparison between the fatigue and alert data on different EEG bands (4 bands) and channels (26 channels).

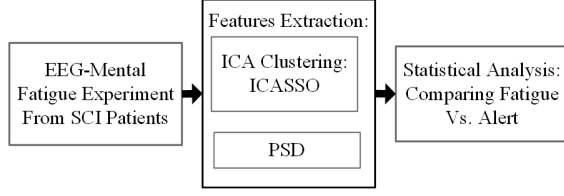


Figure 1. Block diagram of the analysis for SCI mental fatigue study

B. Experiment of Data Collection for Mental Fatigue

This mental fatigue experiment involves 10 SCI patients and 10 healthy persons for comparison with their ages between 17 and 69 years old. The institutional human research ethics committee has approved the experiment [5]. The participants performed cognitively demanding mental tasks and were assessed over a 2 to 3 hours period of time. The experiment is divided into three tasks: pre-task, series of mental workload tasks for inducing mental fatigue and post-task. The EEG measurements were collected on the pre-task and post-task sections. The details of experiment protocols are as follows: (i) For pre-task, participants performed eyes open and closed actions for a duration of 30 seconds each with the EEG measurement. This is used for as an alert data. (ii) For mental workload tasks, participants conducted a series of tasks which were used for inducing mental fatigue including auditory habituation, auditory oddball, stroop, eye tracking, visual working memory, executive maze and prepulse inhibition. The auditory habituation task is used for measuring brain-body response to stimuli. The auditory oddball task is used for assessing basic sensory-motor and decision making. The stroop task is used for testing the executive function. Eye tracking task is for testing eye movement and brain function during tracking. Visual working memory task is used for assessing early language processing and working memory. Executive maze task is used for assessing cognitive function. The prepulse inhibition task corresponds with the startle responses. (iii) For post-task, after finishing the series of mental workload tasks, participants performed another eyes open and closed action for 30 seconds duration with the EEG measurement, which is used for as a fatigue data.

Figure 2 shows the raw EEG data and 26 EEG channels, Quick cap-Neuroscan from Compumedics were used in the experiment. The 26 EEG channels location are: Frontal channels (FP1, FP2, F7, F3, Fz, F4, F8, FC3, FCz and FC4), central channels (C3, Cz, C4, CP3, CPz, CP4), parietal channels (P3, Pz and P4), temporal channels (T3, T4, T5, T6) and occipital channels (O1, Oz, and O2). For reference channels the average of both earlobes at A1 and A2 was

used. The electrode impedance was kept at 5KOhm and below with the sampling rate of 500Hz.

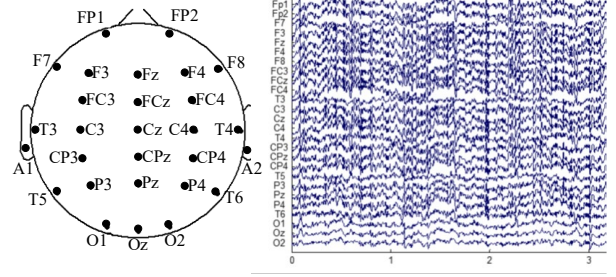


Figure 2. The location and raw data of 26-EEG channels for this study

C. Clustering-based ICA using ICASSO and Power Spectral Density

The features were extracted using ICASSO [4, 6] and power spectral density (PSD) algorithms. ICASSO is a clustering-based ICA which performs ICA several times and the ICA reliability analysis with the following steps: (i) In this study, FastICA was used for the selection of the parameter ICA estimation. The standard ICA model as follows:

$$x = As \quad (1)$$

$$s = Wx \quad (2)$$

where A is the $n \times m$ of the mixing matrix of the ICA model, x is the raw data of 26 EEG channels for this study and s is the independent sources, W is the de-mixing/un-mixing matrix. (ii) The FastICA algorithm above is performed M times to N samples of EEG data with each time, the data is bootstrapped. Here, the de-mixing matrix from the FastICA is calculated repeatedly and by changing of the starting points of the signal at each time. The collection of de-mixing matrix (\hat{W}) is as follows:

$$\hat{W} = [\hat{W}_1^T \hat{W}_2^T \dots \hat{W}_M^T]^T \quad (3)$$

(iii) The results of estimated components are clustered based on their mutual similarity criterion as follows:

$$\sigma_{ij} = |r_{ij}| \quad (4)$$

where σ_{ij} is the final similarity matrix and r_{ij} is the mutual correlation coefficient. (iv) The reliability of the cluster is retrieved and the estimated independent component which in this study used as an input for PSD feature extractor is as follows:

$$I_q(C_m) = \frac{1}{|C_m|^2} \sum_{i,j \in C_m} \sigma_{ij} - \frac{1}{|C_m||C_{-m}|} \sum_{i \in C_m} \sum_{j \in C_{-m}} \sigma_{ij} \quad (5)$$

where I_q denotes the cluster quality index, C_m denotes the set of indices which belongs to m -th cluster, C_{-m} denotes the set of indices that do not belong to the m -th cluster. Note, $I_q(C_m) = 1$ for ideal cluster when (4) is used to calculate the similarity σ_{ij} , and decreases when C_m becomes less

compact and isolated. The estimated independent component (sources) from the ICASSO is used for the data input of the PSD feature extractor. PSD is applied to convert the time domain of the estimated independent component into the frequency domain.

Statistical analysis is used to evaluate and to compare between fatigue and alert states which includes R^2 as the coefficient determination and p -value as the statistical significance. The coefficient determination has the value between 0 and 1 which is referring to the percentage of the variation of the dependent values of the first variable that can explained by the variation of independent values of the second variable. R^2 of 1 is referring of the variation of the first variable being explained fully (100%) by the variation in the second variable [7].

III. RESULTS

The collected EEG dataset from 10 SCI patients is divided into two states, pre-task for alert state and post-task for fatigue state with 30 seconds of data for each experiment and sampling rate of 500 Hz. This results in a matrix with dimension $26 \times 15000 \times 10$ (EEG channels number \times 30second of data with 500Hz sampling rate \times participants number). The experiment is done with the eyes open state and eyes closed state. This paper only uses the eyes open dataset for further processing.

The data from 10 healthy participant is used for comparison purpose. The matrix (26×15000) from each participant is fed to the ICA clustering, ICASSO to cluster the data into 26 clusters based on their similarity. The independent components (sources) resulted from ICASSO are fed to the PSD to convert the time domain signal into the frequency domain. Four EEG bands are calculated from the PSD, including delta (0-3Hz), theta (4-7Hz), alpha (8-13Hz) and beta (14-30Hz). Statistical analysis R^2 and p -value using student t-test are used to compare between the data from alert state and fatigue state on each of the 26 EEG channels and each of the 4 EEG bands.

A. Comparison between SCI and Healthy Groups

TABLE I. COMPARISON BETWEEN SCI AND HEALTHY PARTICIPANTS

Comparison	Statistical significance of p -value			
	Delta	Theta	Alpha	Beta
SCI vs. Healthy on Alert data	0.64	0.46	0.44	0.39
SCI vs. Healthy on Fatigue data	0.45	0.61	0.58	0.49

The statistical analysis is used to compare between SCI group and healthy group on two states of the data (alert and fatigue data) and four EEG bands as shown in Table 1. The p -values as the statistical significance evaluation for alert and fatigue data of SCI patients vs. Healthy participant show no significance difference with values between 0.39 and 0.64 across four EEG bands. With the p -value above 0.05, this shows that data of alert and fatigue from the SCI group are not significantly different or they are comparable with the data from the Healthy group.

B. Further Analysis and Result on SCI Group.

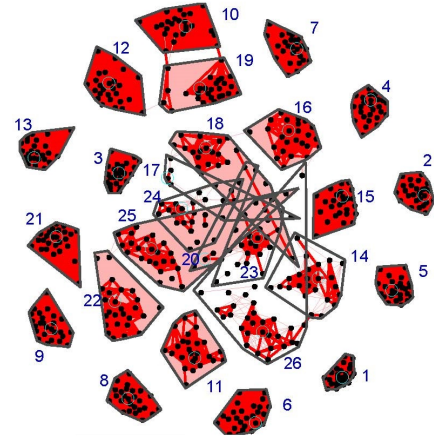


Figure 3. The cluster similarity graph for 'alert' data of SCI group

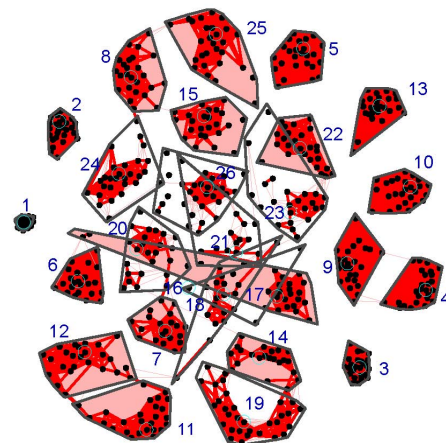


Figure 4. The cluster similarity graph for 'fatigue' data of SCI group

The result of the clustering ICA using ICASSO provided 26 EEG channels. The cluster similarity graphs for alert and fatigue states of SCI group are shown in Fig.3 (alert) and Fig.4 (fatigue). The graphs show a distinct cluster and its relationship (similarity) to other clusters between alert state and fatigue state. For alert clustering (Fig.3), it shows three cluster groups including (i) independent single cluster that has no relationship to other clusters with cluster numbers: from 1 to 9, 11, 12, 13, 21 and 22. (ii) Two clusters show a relationship with the cluster numbers: 10 and 19. The rest of the clusters including cluster from 14 to 18, 20 and clusters from 23 to 26) share a similarity together. For fatigue cluster (Fig. 4), it shows a different similarity to cluster groups including (i) single cluster of 1, 2, 3, 5, 10, 13; (2) Two clusters grouped together including clusters: 4 and 9, 11 and 12, 14 and 19; (iii) The rest of the cluster numbers are grouped together sharing their similarity including clusters 6, 7, 8, cluster from 15 to 18 and clusters from 20 to 26.

Moreover, Fig. 5 shows the coefficient determination (R^2) of fatigue versus alert data from SCI group for proposed clustering ICA method of feature extractor (ICASSO and PSD) and the comparison to standard PSD method across 4 EEG bands and 26 EEG channels. It can be seen that the R^2

values of the proposed features extractor of clustering method (ICASSO & PSD) have lower values compared to the standard PSD across the four EEG bands. In details, for delta band, ICASSO & PSD method has a lower mean R^2 of 0.07 for determining variation of fatigue vs. alert compared to the standard PSD method with mean R^2 of 0.5. For theta and alpha bands, ICASSO & PSD method has a lower mean R^2 of 0.1 for determining variation of fatigue vs. alert compared to the standard PSD method with mean R^2 of 0.6. For beta band, ICASSO & PSD method has also a lower mean R^2 of 0.1 for determining variation of fatigue vs. alert compared to the standard PSD method with mean R^2 of 0.6. Lower R^2 value shows that the proposed clustering ICA method has less correlation and is able to distinguish between each other for comparing fatigue and alert data from SCI patient.

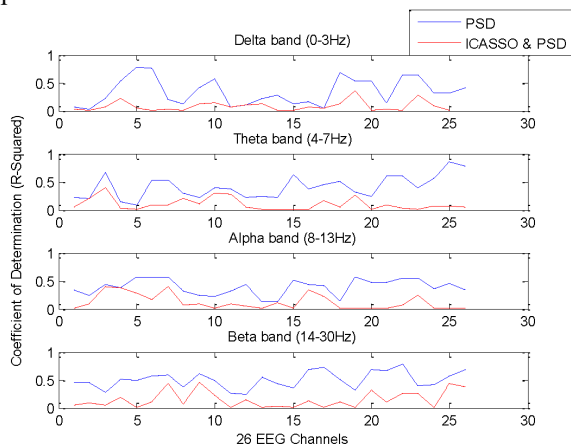


Figure 5. Statistical Analysis for comparing fatigue and alert data using ICASSO & PSD to the PSD alone of SCI group

TABLE II. P -VALUE OF SIGNIFICANCE EEG CHANNELS ALERT VS FATIGUE ACROSS 4 EEG BANDS OF SCI PATIENTS.

Site	Bands	Significance Channels	p -value
Frontal	Delta	F3, Fz, F4, FCz	0.01
	Theta	F7, F8	0.01
	Alpha	FP1, F8	0.02
	Beta	FP1, FP2, F3, Fz, F4, F8, FC3, FCz, FC4	0.02
Central	Delta	No significant channels	>0.05
	Theta	CP3, CPz, CP4	0.03
	Alpha	C3, CP3	0.04
	Beta	Cz, C4	0.02
Parietal	Delta	P4	0.004
	Theta	Pz, P4	0.008
	Alpha	P3, Pz	0.02
	Beta	P3, Pz, P4	0.002
Temporal	Delta	No significant channels	>0.05
	Theta	T4, T6	0.02
	Alpha	T4, T5, T6	0.01
	Beta	T6	0.05
Occipital	Delta	No significance channels	>0.05
	Theta	O1, Oz, O2	0.004
	Alpha	Oz	0.03
	Beta	O1, Oz, O2	0.02

The statistical significance alert vs. fatigue states of the proposed clustering method from SCI group across 4 EEG bands and across 26 EEG channels is given in Table 2. The significant channels with the p -value < 0.05 that influence delta band are frontal channels (F3, Fz, F4 and FCz) and parietal channel (P4). For theta band, the significant channels

with p -value < 0.05 are frontal channels (F7 and F8), central channels (CP3, CPz and CP4), parietal channels (Pz and P4), temporal channels (T4 and T6) and occipital channels (O1, Oz and O2). For alpha band, the significant channels with p -value < 0.05 are frontal channels (FP1, F8), central channels (C3 and CP3), parietal channels (P3 and Pz), temporal channels (T4, T5 and T6) and occipital channel (Oz). For beta band, the significant channels with p -value < 0.05 are frontal channels (FP1, FP2, F3, Fz, F4, F8, FC3, FCz and FC4), central channels (Cz and C4), parietal channels (P3, Pz and P4), temporal channel (T6) and occipital channels (O1, Oz and O2).

IV. CONCLUSION

In this study, the unique combination of clustering ICA (ICASSO) and PCA is used as a feature extraction strategy of the alert and fatigue data from the mental load experiment of ten SCI patients with the EEG measurement. Ten healthy participant are also used for initial comparison between SCI and healthy groups. The statistical analysis (R^2 and p -value) is used for performance measurement. The results show that the comparison SCI vs. healthy group on alert or fatigue data has no statistical significance (p -value > 0.05) which means that they are comparable. Further, the R^2 is calculated from data alert vs. fatigue on SCI patients across 4 EEG bands and 26 EEG channels which results in a lower R^2 compared to using a standard PSD only feature extractor. This means that using the proposed method (ICASSO and PSD), the alert vs. fatigue testing shows more distinction compared to only using PSD method. The statistical significance with p -values less than 0.05 are found across channels and four EEG bands. There are no significant channels found on delta band for central, temporal and occipital locations.

REFERENCES

- [1] A. Craig, Y. Tran, N. Wijesuriya, and J. Middleton, "Fatigue and tiredness in people with spinal cord injury," *Journal of Psychosomatic Research*, vol. 73, pp. 205-210, 2012.
- [2] R. Chai, Y. Tran, G. R. Naik, T. N. Nguyen, S. H. Ling, A. Craig, and H. T. Nguyen, "Classification of EEG based-mental fatigue using principal component analysis and Bayesian neural network," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 4654-4657.
- [3] R. Chai, G. R. Naik, T. N. Nguyen, S. H. Ling, Y. Tran, A. Craig, and H. T. Nguyen, "Driver Fatigue Classification With Independent Component by Entropy Rate Bound Minimization Analysis in an EEG-Based System," *IEEE J. Biomed. Health Informat.*, vol. 21, pp. 715-724, 2017.
- [4] J. Himberg, A. Hyvärinen, and F. Esposito, "Validating the independent components of neuroimaging time series via clustering and visualization," *Neuroimage*, vol. 22, pp. 1214-1222, 2004.
- [5] A. Craig, Y. Tran, N. Wijesuriya, R. Thurasingham, and H. Nguyen, "Switching rate changes associated with mental fatigue for assistive technologies," in *Proc. IEEE 33th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2011, pp. 3071-3074.
- [6] G. R. Naik, A. H. Al-Timemy, and H. T. Nguyen, "Transradial Amputee Gesture Classification Using an Optimal Number of sEMG Sensors: An Approach Using ICA Clustering," *IEEE Trans. Neural. Syst. Rehabil. Eng.*, vol. 24, pp. 837-846, 2016.
- [7] J. Miles, "R squared, adjusted R squared," *Wiley StatsRef: Statistics Reference Online*, 2014.