This is a peer-reviewed, accepted author manuscript of the following conference paper: Kingphai, K & Moshfeghi, Y 2021, 'Mental workload prediction level from EEG signals using deep learning models', Paper presented at The 3rd Neuroergonomics Conference 2021, 11/09/21 - 16/09/21.

Mental Workload Prediction Level from EEG Signals using Deep learning Models

Kunjira Kingphai* **and Yashar Moshfeghi***

NeuraSearch Laboratory, Department of Computer and Information Sciences, University of Strathclyde, Glasgow, Scotland

Correspondence*:
kunjira.kingphai@strath.ac.uk

yashar.moshfeghi@strath.ac.uk

ABSTRACT

Introduction This paper describes our research groups' efforts in tackling the mental workload (MWL) prediction task from Electroencephalogram (EEG) signals which was organised as a Passive BCI Hackathon grand challenge[1] at Neuroergonomics conference 2021. This challenge focuses on the Multi-Attribute Task Battery-II (MATB-II), which has been used for assessing subject MWL capacities; since the MWL has been realised as an essential factor in subject performance within a complicated working system. However, decoding MWL levels (i.e. easy, medium and difficult) from EEG signals is a difficult task. Especially when the training and testing sets are recorded on different sessions, the distribution of the dataset could vary across them (Yin and Zhang, 2017). Thus, this competition was designed to develop algorithms to classify the three MWL levels from EEG signals in an unseen Session. This paper explores the potential of deep learning models to tackle this challenge. The deep learning models explored are Gated Recurrent Unit (GRU), Bidirectional GRU (BGRU), BGRU-GRU, Stacked Long-Short Term Memory (LSTM), Bidirectional LSTM (BLSTM) and BLSTM-LSTM.

Methodology As part of the challenge, EEG signals across three Sessions for each subject were provided, with the signals in Session 1 and 2 being labelled and the ones in Session 3 not being labelled. The EEG signals were collected from 15 subjects (6 female; 9 average 25 y.o.) via 62 electrodes, placed according to the international 10-20 system, sampled at 500 Hz. The participants were invited to the lab for three independent experimental Sessions, each spaced seven days apart. The signals have been recorded in two states; the resting state and the testing state. During resting state, subjects rested with their eyes open for one minute, and their EEG was recorded. Then, subjects completed a MATB-II task with three 5-minute blocks in the testing state, each of a different workload level presented in a pseudo-random manner. To prepare the data for the challenge, the following preprocessing steps were applied to the data by the organisers: epoching, high-pass and low-pass filtering by using FIR filter, referencing and electrode rejection. Moreover, they also down-sampled signal to 250 Hz. In addition to the steps above, we conducted an automatic independent component analysis based on ADJUST (ICA-ADJUST) (Mognon et al., 2011) to remove the artefact components that we observed in the data.

Feature Engineering Since EEG data was provided with a sampling frequency of 250 Hz and epoching into 2-second non-overlapping epochs, we, therefore, calculated the features with

[1] https://www.neuroergonomicsconference.um.ifi.lmu.de/pbci/

a 2-second sliding window size with a 2-second shift. So, we have 500 samples in each 2-second window corresponding to 250 Hz for feature computation. For each window, we have calculated a set of features that can be broadly classified into six groups: frequency, statistical, morphological, time-frequency, linear, and non-linear features. In frequency domain features, we computed signal power for each channel at four well-known Power Spectral Density (PSD) bands which are delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz) and gamma (30–40 Hz). We also computed signal power features in every non-overlapping 2-Hz interval from 4–40 Hz because the non-overlapping 2-Hz could provide finer power spectrum information. While statistical features of mean, variance, skewness, and kurtosis were computed to identify the distribution of signal, three morphological features of curve length, the number of peaks and average non-linear energy were extracted.[2] Wavelet transform was computed to perform time-frequency analysis. The Autoregressive coefficient (AR) with p = 2 was calculated in the linear domain, whereas approximate entropy (ApEn) and Hurst exponent (H) are treated as non-linear features. This has led us to have 3,843 features.[3] We then performed feature selection to optimise the feature set using Support Vector Machine analysis (SVM). The selection was performed in each subject and each session separately, using 10-fold cross-validation. We then performed feature standardisation on the selected features (Buscher et al., 2012). Using the final set of features engineered from Session 1 and 2, we then trained our deep learning models and used that to predict the MWL levels from EEG signals on Session 3. We trained our models using 10-fold cross-validation.

Results Table 1 shows the averaged accuracy of our models trained on Session 1 and tested on Session 2 and vice versa. Our findings show that our models can predict the MWL levels with an accuracy higher than 90%. The overall best model is BGRU-GRU, with an accuracy of 94.023% (SD = 2.074). We also observed that the best performing model for each subject varies, and thus, for this competition, we used that for our submission.

Table 1. The average accuracy and standard deviation of the deep learning models' prediction of the easy, medium and difficult levels of the MWL task.

Stacked GRU	BGRU	BGRU-GRU	Stacked LSTM	BLSTM	BLSTM-LSTM
93.869 (2.785)	93.688 (3.107)	94.023 (2.704)	93.891 (2.785)	93.543 (3.274)	93.840 (2.883)

Summary In this paper, we have described our group's effort in participating in a Passive BCI Hackathon grand challenge at the Neuroergonomics conference 2021. To do so, we have conducted an elaborated feature engineering and used those features in our deep learning models for the tasks of intra-subject MWL estimation based on inter-session adaptation. Overall, all our models have achieved an accuracy of higher than 90% with the best of being 94.023%. Participation in this challenge has inspired us to study further the effect of feature engineering on the MWL level prediction from EEG signals.

Keywords: EEG, Deep Learning, Mental Workload

[2] In this experiment, the four statistical features and three morphological features were extracted in two different ways. Firstly, they were calculated at four frequency bands theta (4–8 Hz), alpha (8–13 Hz), beta (13–25 Hz), and low gamma (25–40 Hz). Secondly, they were extracted from 1-40 Hz as in Chakladar et al. (2020).

[3] The total number of features of a 61-channel data epoch is $63 \times 61 = 3{,}843$.

REFERENCES

Buscher, G., Dengel, A., Biedert, R., and Elst, L. V. (2012). Attentive documents: Eye tracking as implicit feedback for information retrieval and beyond. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 1, 1–30

Chakladar, D. D., Dey, S., Roy, P. P., and Dogra, D. P. (2020). Eeg-based mental workload estimation using deep blstm-lstm network and evolutionary algorithm. *Biomedical Signal Processing and Control* 60, 101989

Mognon, A., Jovicich, J., Bruzzone, L., and Buiatti, M. (2011). Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 48, 229–240

Yin, Z. and Zhang, J. (2017). Cross-session classification of mental workload levels using eeg and an adaptive deep learning model. *Biomedical Signal Processing and Control* 33, 30–47