EEG Mental Workload Classification with Random Forest Classifier

Frank Bolton[1], Dov Te'eni[2], Eran Toch[1]

[1]Department of Industrial Engineering, Tel Aviv University, Tel Aviv, Israel

[2]Coller School of Management, Tel Aviv University, Tel Aviv, Israel

Previous studies have shown that EEG data can be used to distinguish between different levels of mental load, and specifically between the cognitive effort applied in performing a task (Grimes et al., 2008; Dirican and Göktürk, 2011). Passive Brain Computer Interfaces (BCI) use machine learning predictions of mental load made from short durations of EEG recordings in order to change how computer systems operate. This is different than traditional EEG studies that average EEG recordings into grand means, which are then evaluated using descriptive and inferential statistics. Nevertheless, EEG data is still posing many challenges such as being highly non-stationary and having large signal differences between recording sessions and between participants (Fairclough and Lotte, 2020). These challenges make passive BCI systems fragile. However, Passive BCI has been successfully applied to useful tasks like estimating cooperation level in aviation (Verdiere et al., 2019) and predicting the decision that will be made by the user when notifications are received (Bolton et al., 2021). To improve the state of the art in Passive BCI, a standardized EEG dataset was shared openly (Hinss et al., 2021) and a competition held with the Neuroergonomics conference aims to improve performance of passive BCI systems.

The objective of this analysis is to produce a machine learning classifier with the best possible average cross-session intra-participant accuracy at detecting the mental load state from the three possibilities of 'difficult', 'medium' or 'easy'. We documented our method of replicating models described in prior research and then applying general time-series models and refining our design to achieve a best model.

The EEG dataset consists of 15 participants (6 females, average age 25 y.o) who performed three independent experimental sessions spaced 7 days apart. Each session consisted of three 5-minute blocks, each with a different workload level, presented in a random order. Preprocessing performed by the dataset creators produced 149 2-second epochs for each difficulty level. The data consisted of 61-channels and was downsampled to 250Hz. The dataset was downloaded to a local computer the baseline code was reimplemented, and extended with functionality to train a model with all of the labelled data to predict the class of the unlabelled test data. We selected a set of processing and modelling strategies from literature reviews of Deep learning applied to EEG (Roy et al., 2019) and of EEG signal processing (Lotte, 2014).

We used the time-series library sktime (Löning et al., 2019) within a SKLearn pipeline to perform data transformation and classification. A series of Random Forest models with different numbers of model features were trained on the concatenated data from all 61 EEG electrodes, accuracy and time taken are shown in Table 1. We found that 300 features performed best and that a model trained on the first labeled session and tested on the second labeled session produced an average classification accuracy of 50.28% (Figure 1). The model evaluation trained each configuration once using the full set of recordings from the first session and evaluated the accuracy on the full set of recordings from the second session. To produce the submission, the same model was trained on data from the two sessions and predictions made on the unlabeled data (Bolton, 2021). Our contribution to the state of the art is the application of Random Forest to cross-session Passive BCI producing good accuracy.

Number of features	Mean cross-session accuracy	Time (training and evaluation)
80	49.43%	2h06m
200	49.95%	5h09m
300	50.28%	7h42m
400	50.00%	10h15m

Table 1. Accuracy and Time taken to train and test model with different number of Random Forest features. We see a diminishing accuracy with more than 300 features

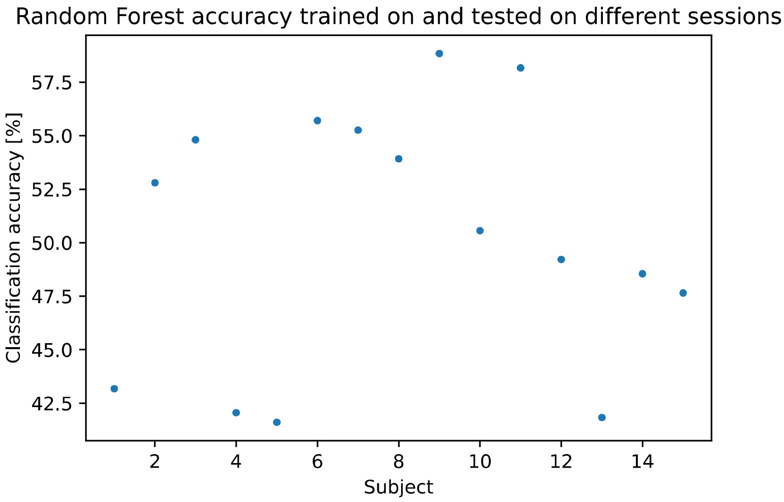

Figure 1. Cross-session classification accuracies (Note our minimum accuracy produced is 41.6%).

References:

Bolton, F. (2021). Bolton PassiveBCIHackathon2021. *GitHub repository*. Available at: https://github.com/frankbolton/PassiveBCIHackathon2021.

Bolton, F., Te'Eni, D., Maimon, N. B., and Toch, E. (2021). Detecting interruption events using EEG. in *2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech)* (IEEE), 33–34.

Christ, M., Braun, N., Neuffer, J., and Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing* 307, 72–77.

Dirican, A. C., and Göktürk, M. (2011). Psychophysiological measures of human cognitive states applied in human computer interaction. *Procedia Computer Science* 3, 1361–1367.

Fairclough, S. H., and Lotte, F. (2020). Grand Challenges in Neurotechnology and System Neuroergonomics. *Front. Neuroergonomics* 1, 602504. doi:10.3389/fnrgo.2020.602504.

Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., and Rao, R. P. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM), 835–844.

Hinss, M. F., Darmet, L., Somon, B., Jahanpour, E., Lotte, F., Ladouce, S., et al. (2021). An EEG dataset for cross-session mental workload estimation: Passive BCI competition of the Neuroergonomics Conference 2021. doi:10.5281/zenodo.5055046.

Löning, M., Bagnall, A., Ganesh, S., and Kazakov, V. (2019). sktime: A Unified Interface for Machine Learning with Time Series. *Workshop on Systems for ML at NeurIPS 2019*, 10.

Lotte, F. (2014). "A Tutorial on EEG Signal-processing Techniques for Mental-state Recognition in Brain–Computer Interfaces," in *Guide to Brain-Computer Music Interfacing*, eds. E. R. Miranda and J. Castet (London: Springer London), 133–161. doi:10.1007/978-1-4471-6584-2_7.

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T. H., and Faubert, J. (2019). Deep learning-based electroencephalography analysis: a systematic review. *J. Neural Eng.* 16, 051001. doi:10.1088/1741-2552/ab260c.

Verdiere, K. J., Dehais, F., and Roy, R. N. (2019). Spectral EEG-based classification for operator dyads' workload and cooperation level estimation. in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (Bari, Italy: IEEE), 3919–3924. doi:10.1109/SMC.2019.8913848.