Deep Neural Network modeling using CNN and BiLSTM for online Passive EEG Brain-Computer Interface to classify Mental Workload

Mario De Lorenzo[1,2], Mali Halac[1], Sukethram Sivakumar[1,2], Hasan Ayaz[2]
[1]College of Engineering, Drexel University
[2]School of Biomedical Engineering, Science and Health Systems, Drexel University

1) Background: Passive brain-computer interfaces (BCIs) are a key element for Neuroergonomics due to their ability to reveal covert brain activities (e.g. mental workload, attention). In turn, this information can be used to improve human-machine interaction (HMI) to construct real-time feedback loops (Ayaz and Dehais, 2019; Lotte and Roy, 2019; Aricò et al., 2016; Dehais et al., 2019). Previous studies demonstrated several proof-of-concept applications of this technology by designing systems that adapt to the user's cognitive or affective states (Ewing et al., 2016; Yuksel et al., 2016). Still, one major challenge remains to be addressed—the large within-subject variability and non-stationarity intrinsic to brain signals such as electroencephalography (EEG) signals. It is therefore crucial for a BCI to operate across sessions or days without recalibration.

2) Methods: In this study, we develop a Deep Neural Network for passive BCI for mental workload detection using the EEG data provided by the Passive BCI Hackathon organized by the Neuroergonomics Conference 2021. ***Dataset:*** The EEG data includes recordings from 15 subjects (6 females and 9 males, with an average age of 25 years) each of which includes 3 different sessions separated by a week. Each session includes recordings of EEG activity while participants completed a MATB-II task with three different difficulty levels— "easy", "medium", "difficult" (administered in a previously determined order). In the easy condition, subjects were given the Track and System Monitoring task. In the medium condition, the Resource Management task was added on. Lastly, in the difficult condition, a communication task was added for a total of 4 tasks simultaneously. The first two sessions are used to train the model whereas the third session was used as a test set. Using a 64 active Ag-AgCl Electrode 10-20 system; 62 of the electrodes were used to gather data. The data is preprocessed by using a high-pass filter and Z-score normalization to unify different EEG metrics to a single metric (Robert et al, 2015). The dataset was split randomly in a stratified fashion into two training and validation sets (70% /30%, respectively). The data was fed to the architecture in a real-time fashion across a given time window. Two time windows were tested (0.0625s and 0.25s), where a shorter time window would result in faster BCIs response, in a tradeoff against lower accuracy. ***Models:*** Our proposed model uses 3 layers of Convolutional Neural Network (CNN) which are fully connected to either 2 layers of Long Short-Term Memory (LSTM) or Bidirectional Long Short-Term Memory (BiLSTM) (Zhang et al., 2017). Also, a simpler model (only BiLSTM layers) was used to calculate the weights of each

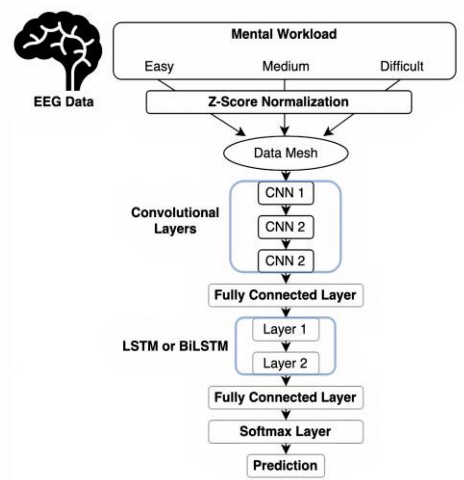

Fig.1 Proposed Deep Neural Network. The tasks are first being uploaded and preprocessed. Z-score normalization is being used to unify different EEG metrics to a single metric. Then, the data is being transposed into a data mesh. The size of the data mesh depends on the number of electrodes being used. The data mesh is fed to the first layers of CNN, then to the layers of LSTM or BiLSTM through a fully connected layer and finally another fully connected layer for prediction.

electrode to filter only the most important electrodes. A further model uses data preprocessed with only 13 prefrontal electrodes (F7, F5, F3, F1, F2, F4, F6, AF3, AFz, AF4, Fp1, Fp2, Fz) since previous studies indicate that the prefrontal cortex reflects changes in mental workload (Geissler et al., 2021). These models were fitted both to the whole dataset and to individual recordings, in order to evaluate inter-patient variability.

3) Results and Discussion: The single-electrode model returned the same accuracy (44%) for each electrode, hence assigning equal weights in predicting the workload. The models that use all electrodes show significantly higher training and validation accuracies. When using all electrodes, the BiLSTM models have a training/validation accuracy of 96%/82%, while LSTM models have a training training/validation accuracy of 96%/81%, When using only the prefrontal electrodes, BiLSTM models have a slightly higher training accuracy than LSTM but a very similar validation accuracy to LSTM when using the same number of epochs. These generic models show a training/validation accuracy 92%/75% hence revealing overfitting. Our findings suggest that decreasing the number of electrodes will slightly decrease the accuracy of generalized models. In addition, when fitting models to single-subject data, maximum training/validation accuracies of BiLSTM were 100%/99% with an average training accuracy of 100% and an average validation accuracy of 92%. The LSTM model performed similarly, with an average of training accuracy of 100% but with a lower validation accuracy of 84%. Interestingly, using 13 electrodes only to create single patient models resulted in lower accuracies (max 99%/89%, average 95%/81%). These results suggest that individual models with BiLSTM layers are more effective compared to generalized models to classify the mental workload while exploiting the data from all 62 electrodes. This may come at a cost in between-patient generality. Overall, our model trained separately for each individual is able to predict the mental workload with a high accuracy overcoming the problem of inter-session variation of EEG signals. We hope to further develop a model that would overcome the inter-subject variation in the future studies.

4) Acknowledgements: We would like to thank Dr. A Duggento and Prof. N Toschi for expert advice and sharing python code (A Duggento) on deep learning models for biological time series. The Titan V GPUs employed in this research were generously donated to N Toschi by NVIDIA.

			Time Window	Dropout	Train Accuracy	Validation Accuracy
13 Electrodes	General Models	LSTM	0.25 s	0.25	0.79	0.73
			0.25 s	0.5	0.89	0.73
			0.25 s	0.25	0.89	0.75
			0.25 s	0.5	0.87	0.75
		BiLSTM	0.25 s	0.25	0.92	0.75
			0.25 s	0.5	0.90	0.74
			0.25 s	0.25	0.91	0.75
			0.25 s	0.5	0.89	0.75
	Subject Models		Time Window	Dropout	Average Train Accuracy	Average Validation Accuracy
		BiLSTM	0.25 s	0.5	0.94	0.8
			0.0625 s	0.5	0.84	0.74
All electrodes	General Models		Time Window	Dropout	Train Accuracy	Validation Accuracy
		LSTM	0.25 s	0.25	0.96	0.81
		BiLSTM	0.25 s	0.25	0.96	0.82
	Subject Models		Time Window	Dropout	Average Train Accuracy	Average Validation Accuracy
		LSTM	0.25 s	0.5	0.99	0.84
		BiLSTM	0.25 s	0.5	0.99	0.92

Table.1 Accuracy results from different models. Different models have been tested to understand for a comparison analysis. The models are divided into models tested with all the electrodes or only the 13 prefrontal electrodes (F7, F5, F3, F1, F2, F4, F6, AF3, AFz, AF4, Fp1, Fp2, Fz), general models which includes all the 15 subjects or subject models individually, different time windows, and dropout. From the accuracies listed, the subject models using all electrodes with BiLSTM layers have a higher train and validation accuracies.

5) References:

Ayaz, Hasan & Dehais, Frédéric. (2019). Neuroergonomics: The Brain at Work and in Everyday Life. https://doi.org/10.1016/C2016-0-02196-4
Lotte, F., & Roy, R. N. (2019). Brain-Computer Interfaces' Contributions to Neuroergonomics. Elsevier. https://doi.org/10.1016/B978-0-12-811926-6.00007-5

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni, F. (2016). A passive brain—Computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. Progress in Brain Research, 228, 295–328. https://doi.org/10.1016/bs.pbr.2016.04.021

Dehais, F., Roy, R. ., & Scannella, S. (2019). Inattentional deafness to auditory alarms: Inter-individual differences, electrophysiological signature and single trial classification. Behavioural Brain Research, 360, 51–59. https://doi.org/10.1016/j.bbr.2018.11.045

Ewing, K. C., Fairclough, S. H., & Gilleade, K. (2016). Evaluation of an Adaptive Game that Uses EEG Measures Validated during the Design Process as Inputs to a Biocybernetic Loop. Frontiers in Human Neuroscience, 10, 223–223. https://doi.org/10.3389/fnhum.2016.00223

Yuksel, B., Oleson, K., Harrison, L., Peck, E., Afergan, D., Chang, R., & Jacob, R. (2016). Learn Piano with BACh: An Adaptive Learning Interface that Adjusts Task Difficulty Based on Brain State. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, 5372–5384. https://doi.org/10.1145/2858036.2858388

Geissler, C.F., Schneider, J. & Frings, C. (2021). Shedding light on the prefrontal correlates of mental workload in simulated driving: a functional near-infrared spectroscopy study. *Sci Rep* **11,** 705. https://doi.org/10.1038/s41598-020-80477-w

Robert W. Thatcher, Carl J. Biver, Duane M. North. (2015).Z Score Neurofeedback. Chapter 1 - History and Technical Foundations of Z Score EEG Biofeedback, 1-21. https://doi.org/10.1016/B978-0-12-801291-8.00001-7

Zhang, Dalin, Lina Yao, Xiang Zhang, Sen Wang, Weitong Chen, and Robert Boots. (2017). Cascade and Parallel Convolutional Recurrent Neural Networks on EEG-Based Intention Recognition for Brain Computer Interface. 32nd AAAI Conference on Artificial Intelligence, AAAI 2018 1703–10. https://arxiv.org/abs/1708.06578v2

Zhang, Dalin & Yao, Lina & Zhang, Xiang & Wang, Sen & Chen, Weitong & Boots, Robert. (2017). EEG-based Intention Recognition from Spatio-Temporal Representations via Cascade and Parallel Convolutional Recurrent Neural Networks. http://arxiv-export-lb.library.cornell.edu/pdf/1708.06578