Graph-Based Learning for EEG Workload Classification: Eliminating the Need for Calibration

Ernesto Bocini
Logitech, Lausanne, 1015 Switzerland
École Polytechnique Fédérale de Lausanne, Lausanne, 1015 Switzerland

Iris Kremer
Logitech, Lausanne, 1015 Switzerland

Pablo Mainar Jovani
Logitech, Lausanne, 1015 Switzerland

Abstract

Overwork and mismanagement of brain workload are leading causes of mental distress. Monitoring Mental Workload (MWL) is therefore crucial for personalized wellbeing recommendations. Electroencephalography (EEG) signals have been shown to assess cognitive states during specific tasks effectively. However, modern methods depend heavily on signal pre-processing and hand-crafted features, limiting their ability to generalize to unseen subjects and often requiring calibration. This study investigates the usage of two graph-based deep learning approaches to tackle these problems. They are tested on two datasets alongside other widely used EEG classifiers. The leave-one-subject-out cross-validation (LOSOCV) strategy is used to tackle the cross-subject generalization problem frequently encountered when using EEG. The results show that models leveraging the graph structure of the EEG data consistently outperform comparison methods on both datasets, achieving strong performance without calibration to new subjects. These results highlight the potential of graph-based approaches as a foundation for future improvements in real-time mental health monitoring and personalized interventions.

Keywords: Mental Workload (MWL); EEG; Graph Neural Networks; LOSOCV

Introduction

Mental Workload (MWL) represents the cognitive effort required to complete a task and plays a crucial role in cognitive performance, productivity and well-being. In recent years, EEG has become a widely used tool for MWL monitoring due to its non-invasive nature, high temporal resolution, and relatively low cost. Traditional MWL classification methods rely on handcrafted EEG features, such as spectral power in specific frequency bands (Brouwer et al., 2012; Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014). To enhance these neuromarkers, machine learning approaches have been explored. These range from traditional classifiers like Support Vector Machines (SVM), Random Forest, and k-Nearest Neighbors (k-NN) (Lim, Sourina, Liu, & Wang, 2015; So, Wong, Mak, & Chan, 2017; Singh & Ahirwal, 2021), to Riemannian Geometry-based models incorporating transfer learning to reduce the need for subject-specific calibration (Kremer, Halimi, Walshe, Cerf, & Mainar, 2024). However, these methods rely on predefined spectral characteristics and usually require calibration for each subject, limiting their generalizability and real-time applicability. More recent deep learning methods aim to eliminate the need for extensive signal pre-processing and manual feature engineering (Sun et al., 2020; Ding, Robinson, Zhang, Zeng, & Guan, 2022; Siddhad, Roy, & Kim, 2025). Graph neural networks (GNNs) have shown promise in EEG-based seizure detection due to their ability to model brain connectivity (Tang et al., 2021; Li, Hwang, Li, Wu, & Ji, 2022), yet their potential for MWL classification has remained underexplored. This work closes this gap by evaluating graph-based models for MWL classification and comparing them to widely used EEG classifiers. To ensure robust, subject-independent evaluation, leave-one-subject-out cross-validation (LOSOCV) is used, which better assesses generalization across individuals.

Methods

Data The evaluation of the models was done on the following two datasets:

- **Mantis** (Fig. 1A.) is a private dataset collected by [Redacted for Anonimity]. EEG data was recorded from 100 subjects (aged 20–71) performing cognitive tasks in a driving simulator. Only N-back (Jonides et al., 1997) tasks were analyzed, using six difficulty levels ($N \in \{0,1,2,3,4,5\}$) as a proxy for MWL. EEG was recorded at 500 Hz with 32 gel-based electrodes (10-20 system).

- **STEW** (Lim, Sourina, & Wang, 2018) is widely used public dataset. EEG data was recorded from 48 subjects performing the SIMKAP multitasking test (Bratfisch & Hagman, 2008), using a 14-channel at 128 Hz (10-20 system). The experiment had two conditions: a 2.5-minute resting baseline and 2.5-minute SIMKAP task.

Figure 1B outlines the preprocessing and analysis pipeline for EEG datasets. Both datasets underwent the same standard preprocessing steps, including resampling to 250 Hz, notch filtering at 50 Hz, bad channel removal, and referencing to a common average. In step 3 (Reshape), the Mantis dataset was split into 1s epochs with an overlap of 0.3, while the STEW dataset used an overlap of 0.5. Additionally, the number of stacked epochs (K) was set to 30 for Mantis and 15 for STEW, accounting for the different amounts of data available.

Classifiers The two main candidates with graph-based approaches are the following:

- **GGN**: The architecture (Li et al., 2022) integrates a connectivity graph generator, spatial decoder with attentive graph convolution, CNNs, and a fully connected classifier. It extracts brain connectivity from EEG data, refines features via CNNs, and classifies seizure types using a fully connected layer.

- **GNN**: The Self-Supervised GNN-SSL (Tang et al., 2021) models EEG as a graph, using electrode geometry or dynamic connectivity to capture spatiotemporal dependencies.

To have a comparison with the most widely used models for MWL, the performance of CNN (Asif, Roy, Tang, & Harrer, 2020), and Transformer (Yan, Li, Xu, Yu, & Xu, 2022; Siddhad, Gupta, Dogra, & Roy, 2024) is also evaluated.

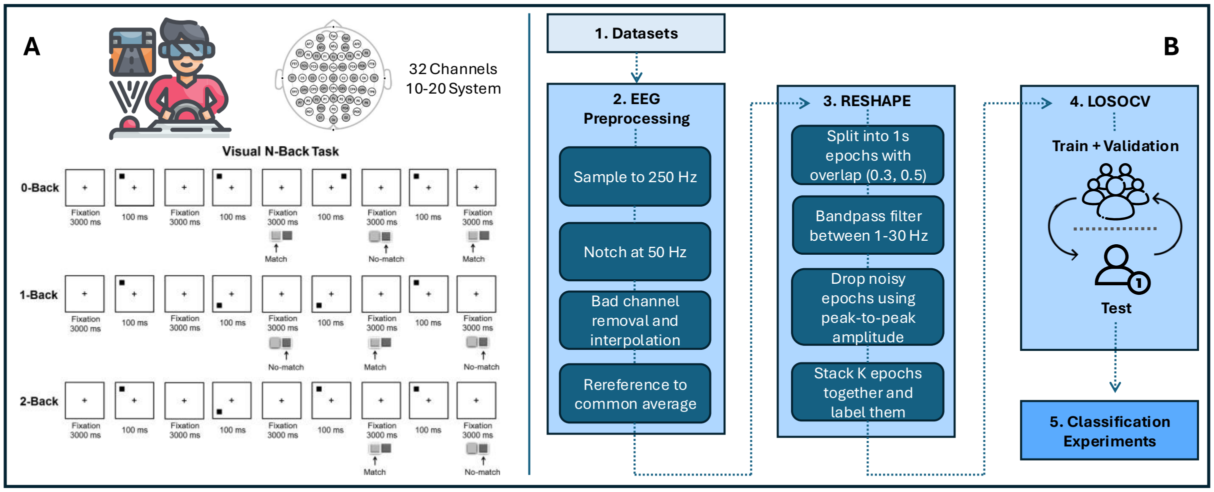

Figure 1: **A:** Mantis dataset: EEG data collected with 32 electrodes placed accordingly to the 10-20 system while involved in driving a simulator and performing N-Back memory task; **B:** End-to-End pipeline.

Table 1: Classification results across different Datasets and tasks. The best-performing model for each task is highlighted in **bold**.

Model	**Mantis 1** Accuracy $\pm$ Std	**Mantis 2** Accuracy $\pm$ Std	**Mantis 3** Accuracy $\pm$ Std	**STEW Dataset** Accuracy $\pm$ Std
GGN	0.83 $\pm$ 0.017	**0.77** $\pm$ 0.012	0.50 $\pm$ 0.050	**0.79** $\pm$ 0.026
GNN	0.73 $\pm$ 0.010	0.70 $\pm$ 0.010	**0.54** $\pm$ 0.060	0.57 $\pm$ 0.014
Transformer	0.79 $\pm$ 0.012	0.62 $\pm$ 0.010	0.47 $\pm$ 0.060	0.51 $\pm$ 0.015
CNN	**0.84** $\pm$ 0.017	0.66 $\pm$ 0.009	0.42 $\pm$ 0.050	0.72 $\pm$ 0.028

Evaluation Strategy MWL literature is usually evaluating their models using a classical train-val-test split after combining and shuffling all the data from all subjects and all sessions (Sun et al., 2020; Ding et al., 2022). Given the high level of inter-subject variability of EEG responses to same type of stimuli,it is important to make sure that good accuracy is maintained when the model is tested on unseen participants (Kingphai & Moshfeghi, 2024). Figure 1B.5 demonstrates the data splitting process: set aside the full data from one subject for testing while the rest is partitioned into train and validation using a standard 80-20 split. This procedure is repeated until each subject is used once as a testing subject.

Results

The classifiers have been evaluated on the following tasks:
Baseline (Mantis 1): classify between eyes open and eyes closed. Easy task, sanity check;
Easy vs Hard (Mantis 2): classify whether the participant is doing N-Back level (N $\in \{0,1\}$) *easy*, or a *hard* (N $\in \{4,5\}$);
Easy vs Medium vs Hard (Mantis 3): classify whether the participant is doing N-Back level *easy* (N $\in \{0,1\}$), *medium* (N $\in \{2,3\}$) or *hard* (N $\in \{4,5\}$;
STEW: classify between *easy* (no task) or *hard*(SIMKAP task).

Table 1 contains results for all these tasks. It can be observed that everywhere except on the Baseline (Mantis 1), the graph approaches are beating previous state-of-the-art results.

Conclusions

This study demonstrates that graph-based models are shown to outperform traditional EEG classifiers for MWL assessment, particularly in their ability to generalize across subjects without requiring calibration data. By leveraging the natural network structure of EEG signals, these models provide a promising path toward more robust, subject-independent MWL monitoring. Future work will focus on expanding dataset diversity, improving interpretability, and integrating real-time applications for adaptive cognitive monitoring.

References

Asif, U., Roy, S., Tang, J., & Harrer, S. (2020). Seizurenet: Multi-spectral deep feature learning for seizure type classification. In *Machine learning in clinical neuroimaging and radiogenomics in neuro-oncology: Third international workshop, mlcn 2020, and second international workshop, rno-ai 2020, held in conjunction with miccai 2020, lima, peru, october 4–8, 2020, proceedings 3* (pp. 77–87).

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, *44*, 58–75.

Bratfisch, O., & Hagman, E. (2008). Simkap–simultankapazität/multi-tasking. *Mödling: Schuhfried GmbH*.

Brouwer, A.-M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using eeg spectral power and erps in the n-back task. *Journal of neural engineering*, *9*(4), 045008.

Ding, Y., Robinson, N., Zhang, S., Zeng, Q., & Guan, C. (2022). Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Transactions on Affective Computing*, *14*(3), 2238–2250.

Jonides, J., Schumacher, E. H., Smith, E. E., Lauber, E. J., Awh, E., Minoshima, S., & Koeppe, R. A. (1997). Verbal working memory load affects regional brain activation as measured by pet. *Journal of cognitive neuroscience*, *9*(4), 462–475.

Kingphai, K., & Moshfeghi, Y. (2024). Mental workload assessment using deep learning models from eeg signals: a systematic review. *IEEE Transactions on Cognitive and Developmental Systems*.

Kremer, I., Halimi, W., Walshe, A., Cerf, M., & Mainar, P. (2024). Predicting cognitive load with eeg using riemannian geometry-based features. *Journal of Neural Engineering*, *21*(5), 056002.

Li, Z., Hwang, K., Li, K., Wu, J., & Ji, T. (2022). Graph-generative neural network for eeg-based epileptic seizure detection via discovery of dynamic brain functional connectivity. *Scientific reports*, *12*(1), 18998.

Lim, W. L., Sourina, O., Liu, Y., & Wang, L. (2015). Eeg-based mental workload recognition related to multitasking. In *2015 10th international conference on information, communications and signal processing (icics)* (pp. 1–4).

Lim, W. L., Sourina, O., & Wang, L. P. (2018). Stew: Simultaneous task eeg workload data set. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(11), 2106–2114.

Siddhad, G., Gupta, A., Dogra, D. P., & Roy, P. P. (2024). Efficacy of transformer networks for classification of eeg data. *Biomedical Signal Processing and Control*, *87*, 105488.

Siddhad, G., Roy, P. P., & Kim, B.-G. (2025). Neural networks meet neural activity: Utilizing eeg for mental workload estimation. In *International conference on pattern recognition* (pp. 325–339).

Singh, U., & Ahirwal, M. K. (2021). Mental workload classification for multitasking test using electroencephalogram signal. In *2021 ieee international conference on technology, research, and innovation for betterment of society (tribes)* (pp. 1–6).

So, W. K., Wong, S. W., Mak, J. N., & Chan, R. H. (2017). An evaluation of mental workload with frontal eeg. *PloS one*, *12*(4), e0174949.

Sun, Z., Li, B., Duan, F., Jia, H., Wang, S., Liu, Y., ... Sole-Casals, J. (2020). Wlnet: Towards an approach for robust workload estimation based on shallow neural networks. *Ieee Access*, *9*, 3165–3173.

Tang, S., Dunnmon, J. A., Saab, K., Zhang, X., Huang, Q., Dubost, F., ... Lee-Messer, C. (2021). Self-supervised graph neural networks for improved electroencephalographic seizure analysis. *arXiv preprint arXiv:2104.08336*.

Yan, J., Li, J., Xu, H., Yu, Y., & Xu, T. (2022). Seizure prediction based on transformer using scalp electroencephalogram. *Applied Sciences*, *12*(9), 4158.