The Journal of Open Source Software

SelfEEG: A Python library for Self-Supervised Learning in Electroencephalography

Federico Del Pup [1,2,3], **Andrea Zanola** [3*], **Louis Fabrice Tshimanga** [2,3*], **Paolo Emilio Mazzon**[3], and **Manfredo Atzori** [2,3,4]

1 Department of Information Engineering, University of Padova, Via Gradenigo 6/b, 35131 Padova, Italy **2** Department of Neuroscience, University of Padua, Via Belzoni 160, 35121 Padova, Italy **3** Padova Neuroscience Center, University of Padova, Via Orus 2/B, 35129 Padova, Italy **4** Information Systems Institute, University of Applied Sciences Western Switzerland (HES-SO Valais), 2800 Sierre, Switzerland

***** These authors contributed equally.

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Elizabeth DuPre

Reviewers:

- @vferat
- @wmvanvliet
- @Bsingstad

Submitted: 16 December 2023
Published: unpublished

Summary

SelfEEG is an open-source Python library developed to assist researchers in conducting Self-Supervised Learning (SSL) experiments on electroencephalography (EEG) data. Its primary objective is to offer a user-friendly but highly customizable environment, enabling users to efficiently design and execute self-supervised learning tasks on EEG data.

SelfEEG covers all the stages of a typical SSL pipeline, ranging from data import to model design and training. It includes modules specifically designed to: split data at various granularity levels (e.g., session-, subject-, or dataset-based splits); effectively manage data stored with different configurations (e.g., file extensions, data types) during mini-batch construction; provide a wide range of standard deep learning models, data augmentations and SSL baseline methods applied to EEG data.

Most of the functionality offered by selfEEG can be executed both on GPUs and CPUs, expanding its usability beyond the self-supervised learning area. Additionally, selfEEG can be employed for the analysis of other biomedical signals often coupled with EEGs, such as electromyography or electrocardiography data.

These features make selfEEG a versatile deep learning tool for biomedical applications and a useful resource in SSL, one of the currently most active fields of Artificial Intelligence.

Statement of need

SelfEEG answers to the lack of Self-Supervised Learning (SSL) frameworks for the analysis of EEG data.

In fact, despite the recent high number of publications (more than 20 journal papers in the last 4 years (Del Pup & Atzori, 2023)), there are currently no frameworks or common standards for developing EEG-based SSL pipelines, contrary to other fields such as computer vision (see LightlySSL or ViSSL).

In the field of EEG data analysis, where it has been demonstrated that SSL can improve models' accuracy and mitigate overfitting (Rafiei et al., 2022) (Banville et al., 2021), the absence of a self-supervised learning framework dedicated to EEG signals limits the development of novel strategies, reproducibility of results, and the progress of the field.

Thanks to selfEEG, researchers can instead easily build SSL pipelines, speeding up experimental design and improving the reproducibility of results. Reproducibility is a key factor in this area, as it enhances the comparison of different strategies and supports the creation of useful benchmarks.

SelfEEG was also developed considering the needs of deep learning researchers, for whom this library has been primarily designed. For this reason, selfEEG aims to preserve a high but easily manageable level of customization.

Library Overview

SelfEEG is a comprehensive library for SSL applications to EEG data. It is built on top of PyTorch (Paszke et al., 2019) and includes several modules targeting all the steps required for developing EEG-based SSL pipelines. In particular, selfEEG comprises the following modules:

- **dataloading**: a collection of functions and classes designed to support data splitting and the construction of efficient PyTorch dataloaders in the EEG context.
- **augmentation**: a collection of EEG data augmentation functions and other classes designed to combine them in more complex patterns.
- **models**: a collection of EEG deep learning models.
- **losses**: a collection of self-supervised learning losses.
- **ssl**: a collection of self-supervised learning algorithms applied to EEG analysis with highly customizable fit methods.
- **utils**: a collection of utility functions and classes for various purposes, such as a PyTorch compatible EEG sampler and scaler.

Related open-source projects

Despite several deep learning frameworks having been developed for the analysis of EEG data, a library focused on the construction of self-supervised learning pipelines on EEG data is still not available to the best of our knowledge, hindering the advancement of the scientific knowledge and the progress in the field. A comprehensive review of open-source projects related to neuroscientific data analysis is provided in (Tshimanga et al., 2023). Few examples are EEG-DL (Hou et al., Feb. 2020) and torchEEG, which characterized for their completeness and spread among the neuroscientific community.

Future development

Considering how rapidly self-supervised learning is evolving, this library is expected to be constantly updated by the authors and the open-source community, especially by adding novel SSL algorithms, deep learning models, and functionalities that can enhance the comparison between different developed strategies. In particular, the authors plan to continue working on selfEEG during the next years via several ongoing European and national projects.

CRediT Authorship Statement

FDP: Conceptualization, Writing - Original Draft, Software - Development, Software - Design, Software - Testing; AZ: Writing - Review & Editing, Software - design (dataloading and utils modules), Software - Testing; LFT: Writing - Review & Editing, Software - design (dataloading and utils modules), Software - Testing; PEM: Technical support, Writing - Review & Editing, Software - Testing; MA: Funding Acquisition, Project Administration, Supervision, Writing - Review & Editing.

Acknowledgements

This work was supported by the STARS@UNIPD funding program of the University of Padova, Italy, through the project: MEDMAX. This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement no 101137074 - HEREDITARY. We would also like to thank the other members of the Padova Neuroscience Center for their support during the project development.

References

Banville, H., Chehab, O., Hyvärinen, A., Engemann, D.-A., & Gramfort, A. (2021). Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering*, *18*(4), 046020. https://doi.org/10.1088/1741-2552/abca18

Del Pup, F., & Atzori, M. (2023). Applications of self-supervised learning to biomedical signals: A survey. *IEEE Access*, *11*, 144180–144203. https://doi.org/10.1109/ACCESS.2023.3344531

Hou, Y., Zhou, L., Jia, S., & Lun, X. (Feb. 2020). A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN. *Journal of Neural Engineering*, *17*(1), 016048. https://doi.org/10.1088/1741-2552/ab4af6

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., & others. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*. https://doi.org/10.48550/arXiv.1912.01703

Rafiei, M. H., Gauthier, L. V., Adeli, H., & Takabi, D. (2022). Self-supervised learning for electroencephalography. *IEEE Transactions on Neural Networks and Learning Systems*. https://doi.org/10.1109/TNNLS.2022.3190448

Tshimanga, L. F., Del Pup, F., Corbetta, M., & Atzori, M. (2023). An overview of open source deep learning-based libraries for neuroscience. *Applied Sciences*, *13*(9). https://doi.org/10.3390/app13095472