The Journal of Open Source Software

Pyheartlib: A Python package for processing electrocardiogram signals

Sadegh Mohammadi [1]

1 Independent Researcher

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

License
Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

Summary

Electrocardiogram (ECG) signals represent the electrical activity of the heart as a graph of voltage versus time. These signals have significant importance in healthcare and contain valuable information. Therefore, they can be analyzed for the diagnosis of various disease.

`Pyheartlib` is a Python package for processing electrocardiogram recordings. This software facilitates working with signals for tasks such as heartbeat detection, heartbeat classification, and arrhythmia classification. Utilizing it, researchers can focus on these tasks without the burden of designing data processing modules. The package transforms original data into processed signal excerpts and their computed features in order to be used for training various machine learning models including advanced deep learning models, which can be trained by taking advantage of Keras (Chollet & others, 2015) and Tensorflow (Abadi et al., 2015) libraries.

Statement of need

Recent advances in deep learning have made it possible to accomplish tasks previously unattainable through the implementation of conventional algorithms. The capabilities of conventional approaches for analyzing ECG signals are limited. Therefore, it is imperative to develop data-driven models capable of automatically identifying patterns from vast heterogeneous data. `Pyheartlib` supports this goal and, unlike popular packages such as NeuroKit2 (Makowski et al., 2021), does not implement pre-defined algorithms. It instead provides data processing modules to assist researchers with the implementation of their own models for various tasks. For example, SleepECG (Brunner & Hofer, 2023) and Neurokit2 (Makowski et al., 2021) have implemented several conventional methods, such as the Pan–Tompkins (Pan & Tompkins, 1985) and Hamilton (Hamilton, 2002) algorithms, for the R-peak detection task. However, the same task can be accomplished through deep learning. Various models can be trained by utilizing the readily available data processing modules in `Pyheartlib`. An example model is provided with this package that performs this task. `Pyheartlib` comprises several dedicated classes for different use cases.

For the heartbeat classification task, which typically requires segmented and annotated heartbeat waveforms, the software preprocesses the input data and delivers datasets comprising waveforms and features. Features can be computed from waveform and RR-intervals.

For the classification of signal excerpts, e.g., arrhythmia classification, the software is designed to store metadata about the excerpts in order to reduce memory usage significantly, especially in situations that each excerpt is being used only once during model training. By adjusting the parameters appropriately, the package provides a substantial quantity of data samples for training deep learning models. Moreover, it is feasible to incorporate RR-intervals in addition to waveforms and their computed features.

DRAFT

Another use case is when each excerpt has to be divided into smaller sub-segments each with a specific label, e.g., R-peak detection. The package delivers data samples by storing metadata about the excerpts and providing lists of labels as annotations for the excerpts.

`Pyheartlib` is easy to use and its documentation contains examples for different use cases. Figure 1 illustrates an example outcome of a deep learning model that was trained with the help of `Pyheartlib` and Keras (Chollet & others, 2015) to detect R-peaks.

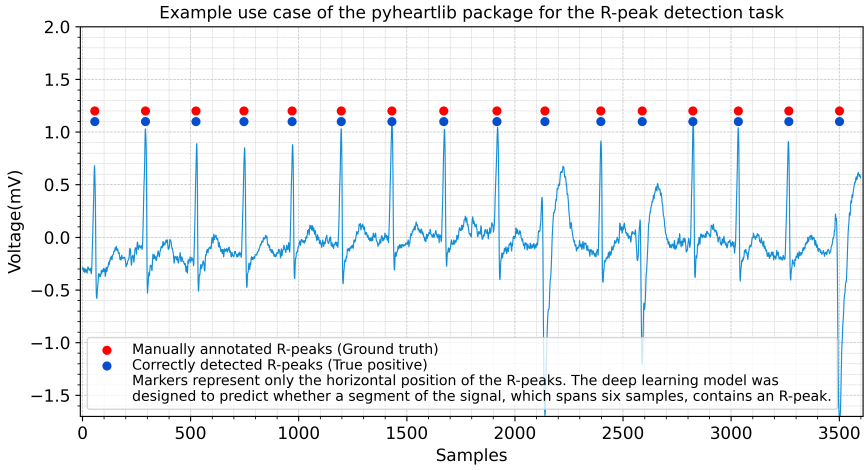

Figure 1: R-peak detection using deep learning. In this example use case, the Pyheartlib package was utilized for processing the signals for training and testing a deep learning model that was implemented using the Keras library. This figure illustrates the outcome of the model on the test dataset. Additional information about this example model is available on the project repository and its documentation.

This package utilizes the WFDB Python package to read data, therefore it supports the WFDB format. Recordings with different formats can be converted using the WFDB Python package (Goldberger et al., 2000; G. Moody et al., 2021; G. B. Moody & Mark, 2001; Xie et al., 2023).

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., … Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*. https://www.tensorflow.org/

Brunner, C., & Hofer, F. (2023). SleepECG: A python package for sleep staging based on heart rate. *Journal of Open Source Software*, *8*(86), 5411. https://doi.org/10.21105/joss.05411

Chollet, F., & others. (2015). *Keras*. https://keras.io

Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., Mietus, J. E., Moody, G. B., Peng, C. K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*, *101*(23), e215–e220. https://doi.org/10.1161/01.CIR.101.23.e215

Hamilton, P. (2002). *Open source ECG analysis*. 101–104. https://doi.org/10.1109/CIC.2002.1166717

Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal

DRAFT

processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Moody, G. B., & Mark, R. G. (2001). The impact of the MIT-BIH arrhythmia database. *IEEE Engineering in Medicine and Biology Magazine*, *20*(3), 45–50. https://doi.org/10.1109/51.932724

Moody, G., Pollard, T., & Moody, B. (2021). WFDB software package (version 10.6.2). In *PhysioNet*. https://doi.org/10.13026/zzpx-h016

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, *BME-32*(3), 230–236. https://doi.org/10.1109/TBME.1985.325532

Xie, C., McCullum, L., Johnson, A., Pollard, T., Gow, B., & Moody, B. (2023). Waveform database software package (WFDB) for python (version 4.1.0). In *PhysioNet*. https://doi.org/10.13026/9njx-6322

DRAFT