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: Britta Westner

Reviewers:

- @cudmore
- @LegrandNico

Submitted: 17 June 2023
Published: unpublished

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 cardiac conditions.

`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

A popular software for working with electrocardiogram recordings is the WFDB Python package, which has the ability to read and write recordings and perform some basic signal processing operations (Goldberger et al., 2000; G. Moody et al., 2021; G. B. Moody & Mark, 2001; Xie et al., 2023). NeuroKit is another Python package with some functionalities such as denoising, delineation, and plotting (Makowski et al., 2021). These packages don't have the ability to prepare electrocardiogram recordings for training machine-learning models for various tasks. `Pyheartlib` addresses this issue by providing researchers with the necessary data processing modules. It 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.

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 heartbeats.

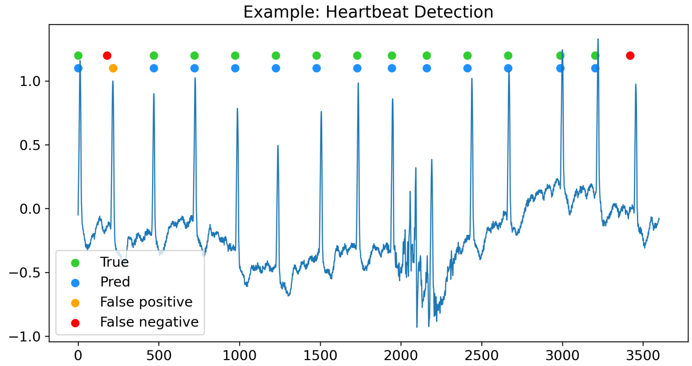

Figure 1: Example: heartbeat detection using deep learning.

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/

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

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 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

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