The Journal of Open Source Software

WearableHRV: A Python package for the validation of heart rate and heart rate variability in wearables

Mohammadamin Sinichi [1,3¶], **Martin Gevonden** [2,3], **and Lydia Krabbendam** [1,3]

1 Department of Clinical, Neuro- & Developmental Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, The Netherlands **2** Department of Biological Psychology, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, The Netherlands **3** Institute Brain and Behaviour (iBBA), Amsterdam, The Netherlands ¶ Corresponding author

DOI: 10.21105/joss.06240

Software

- Review
- Repository
- Archive

Editor: Arfon Smith

Reviewers:

- @AKuederle
- @richrobe

Submitted: 10 November 2023
Published: 23 August 2024

Summary

Wearable devices that monitor physiology have become ubiquitous, and include smart watches, smart jewellery, and smart textiles. The market for these devices is rapidly expanding with new brands and products. These devices measure a variety of signals, which are translated into a large amount of different features, of which heart rate (HR) and heart rate variability (HRV) are among the most common. These features are particularly interesting, not only for consumers but also for researchers, because they are predictive of mental and physical health outcomes, and easily obtained. However, for manufacturers, there may be a trade-off between user acceptability on the one hand and accuracy on the other, where profit margins typically turn out to be the decisive factor. Therefore, the following question continually comes up: is the cardiac data recorded by this new watch/ring/shirt accurate enough to use in research?

The `WearableHRV` Python package offers a comprehensive pipeline for validating the accuracy of HR and HRV measurements. It allows for advanced statistical analyses on device agreement from beat-to-beat cardiac data. The package's graphical user interface (GUI) facilitates pre-processing, visualization, and data analysis at both individual and group levels. As input, a user of `WearableHRV` should use a criterion device (i.e., a gold standard), preferably an electrocardiograph (ECG). Simultaneously acquired data from the device(s) of interest can then be compared against the criterion. The only required inputs for the pipeline are the inter-beat intervals (IBIs) and timestamps for each device; the rest is fully handled by the `WearableHRV` package.

Statement of Need

The use of wearables in psychophysiology and sports sciences has exponentially increased over the past decade. While the golden standard is electrocardiography, for prolonged monitoring many studies use devices that measure HR and HRV using Photoplethysmography (PPG). This technique optically measures the changes in blood volume in peripheral tissues (e.g., earlobe, wrist, arm, fingertip) and substitutes detected pulses for the actual contraction of cardiac ventricular muscles(Challoner & Ramsay, 1974). Although the PPG method is promising and versatile, several studies have highlighted concerns about its validity in identifying HR and HRV, especially when the user is in motion (Allen, 2007; Bent et al., 2020; Hill et al., 2015; Nederend et al., 2017; Pinheiro et al., 2016; Quintana et al., 2016; Schäfer & Vagedes, 2013; Stone et al., 2021).

The oversight regarding the accuracy of these wearables may be due to the lack of an integrated and user-friendly method for assessing the validity of new wearables entering the market. A validation pipeline could lead users from collecting raw data, through pre-processing and advanced statistical analysis, to obtaining the necessary parameters and visualizations required to assess device agreement. The `WearableHRV` package was developed to address this gap. Other existing packages such as `hrv-analysis` (Robin Champseix, 2021), `NeuroKit2` (Makowski et al., 2021), `pyHRV` (Gomes et al., 2019), and similar toolkits offer solutions for pre-processing, analysis, and visualization once IBIs from a single device are provided. However, the added value of `WearableHRV` is rooted in a few key aspects:

First, the currently available packages are not tailored for validation purposes. This becomes especially noticeable when validating multiple wearables at once, across different experimental conditions. This manual process can become a cumbersome task with the packages currently at hand. Second, to make the validation of wearables easier, a user-friendly solution is needed, which current packages lack. This is particularly notable for essential validation steps, such as correcting linear and non-linear lags between devices or trimming signals for a specific device or condition, steps that `WearableHRV` simplifies with GUIs. Third, establishing the validity of a wearable against a gold standard involves numerous decisions. Our pipeline provides a thorough method with extensive documentation for users to do this. The availability of such a pipeline also encourages researchers to contribute to establishing a standardized validation protocol, unifying the approach, reducing variability between methods, and facilitating result comparisons. Finally, the division of the `WearableHRV` pipeline into individual and group pipelines offers an advantage for different types of users. The individual pipeline is designed for processing data from a single participant, whereas the group pipeline offers tools to establish the quality of the signals, device agreement, and validity of the devices across multiple participants. Most common statistical analyses in validation studies, such as mean absolute percentage error, regression analysis, intraclass correlation coefficient (ICC), and Bland-Altman analysis, are already incorporated into the pipeline (Altman & Bland, 1983; Bruton et al., 2000; Haghayegh et al., 2020; Makridakis, 1993).

In summary, provided that a wearable device (either PPG or ECG) allows for the export of the complete time series of recorded IBIs, this Python package makes it relatively easy to establish the validity of a novel wearable in just a few steps. The inclusion of GUI in most functions grants researchers and wearable users the flexibility to validate an unlimited number of wearables across a wide range of conditions.

Main Features and Basic Usage

In this section, we provide an overview of the main functions and basic usage of the `WearableHRV` pipeline. The complete documentation of the API and modules can be found here. Please also refer to the README.md, which provides links to several Jupyter Notebook examples.

The pipeline is divided into two parts: the individual pipeline, which focuses on the data from a single individual, and the group pipeline, which combines the data from all the participants in a single study.

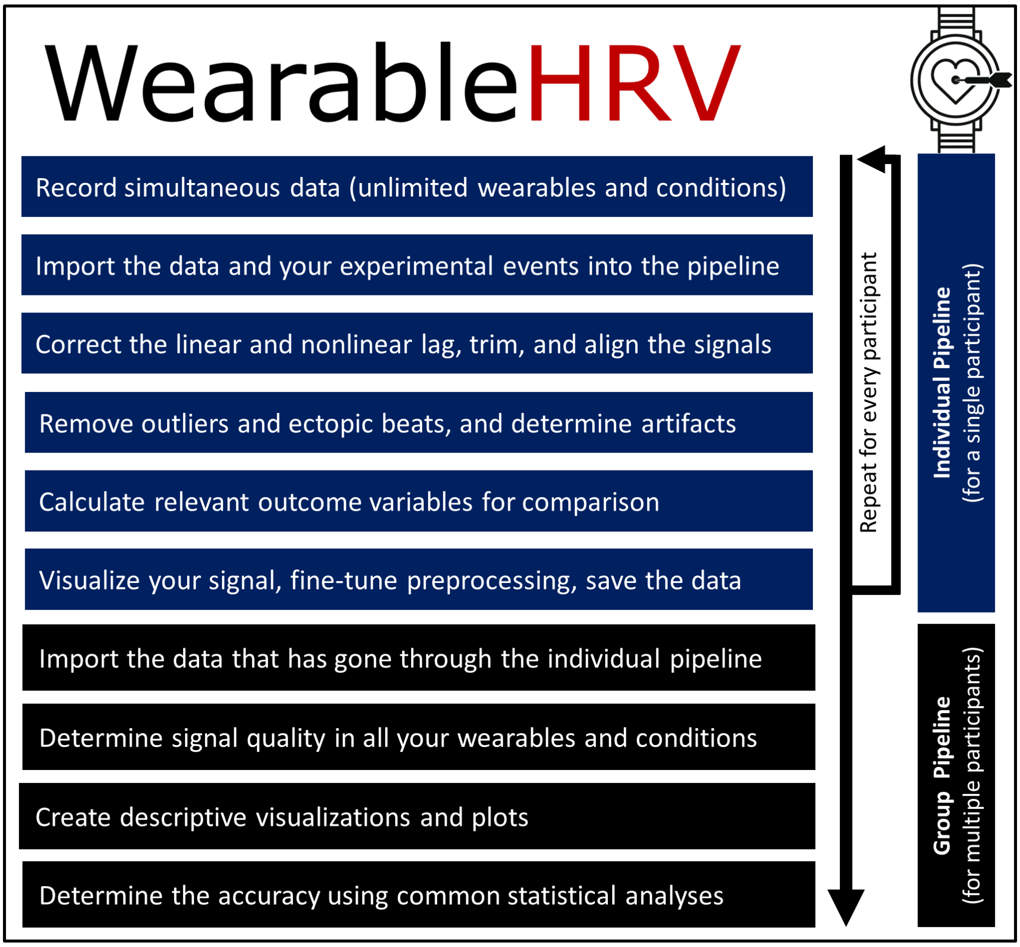

Figure 1: A summary of the main functionalities of WearableHRV Python package.

Individual pipeline

The individual pipeline allows for the validation of a wearable of interest against a criterion device (e.g., a gold-standard ECG) under different conditions. Data necessary for the pipeline should be in a .csv file with UNIX timestamps (specified in milliseconds) and IBIs, in two separate columns. Next, experimental conditions need to be defined using the `define_events` function. The `import_data` function is then used to bring the continuous time series from all devices into the pipeline. Data segmentation into smaller, condition-specific chunks is achieved with the `chop_data` function.

One of the primary strengths of `WearableHRV` is the `visual_inspection` function with the assistance of the GUI, which allows for simultaneous visualization of IBIs and addresses the challenge of correcting devices' lag when wearables' internal clocks are not in sync.

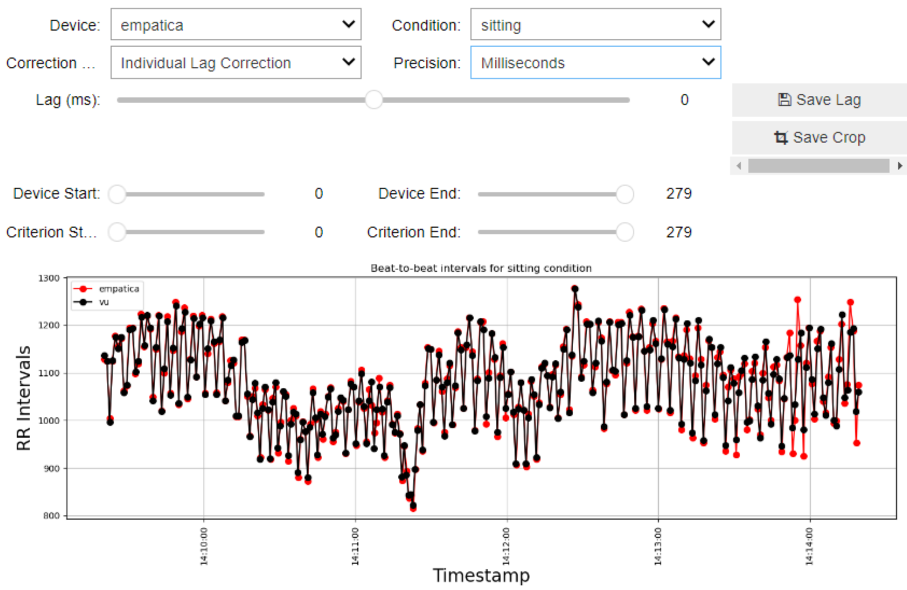

Figure 2: The black line shows the IBIs of the criterion device (ECG). The red line represents the IBIs of a given PPG device. By toggling the `Device` and `Conditions` widgets, one can easily explore other devices and conditions. The `Lag` slider allows for lag correction between the devices, and it is possible to crop a part of the signal if necessary.

Pre-processing and feature extraction with the `pre_processing` and `data_analysis` functions use the `hrv-analysis` python package functionalities (Robin Champseix, 2021), and output numerous time domain and frequency domain features for each condition and device. To facilitate comparing the criterion device with the device of interest and visualization of results, several plotting functions are provided, including `result_comparison_plot`, `bar_plot`, `line_plot`, `radar_plot`, and `unfolding_plot`. All time and frequency features for every device and condition can be exported for later use in the group pipeline via the `save_data` function.

Group pipeline

Moving to group-level analysis, the `import_data` function of the group pipeline aggregates .csv output files that have gone through the individual pipeline.

An important step in wearable validation is quantifying signal quality, which the `signal_quality` function allows for labeling as poor, missing, and acceptable based on specified criteria by users. The results can be visualized through two functions: `signal_quality_plot1` and `signal_quality_plot2`.

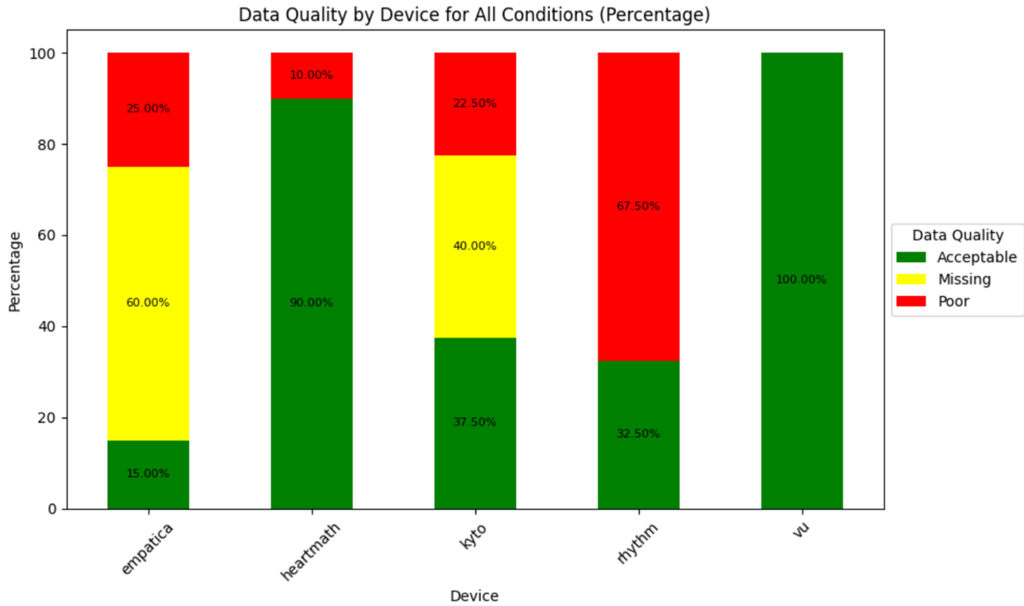

Figure 3: An example of a signal quality plot, achievable by setting a few thresholds. Each bar corresponds to a device, and the y-axis shows the percentage of poor, acceptable, and missing data in each of them.

For visualization across the aggregated dataset, the group module offers `violin_plot`, `box_plot`, `radar_plot`, `hist_plot`, and `matrix_plot`.

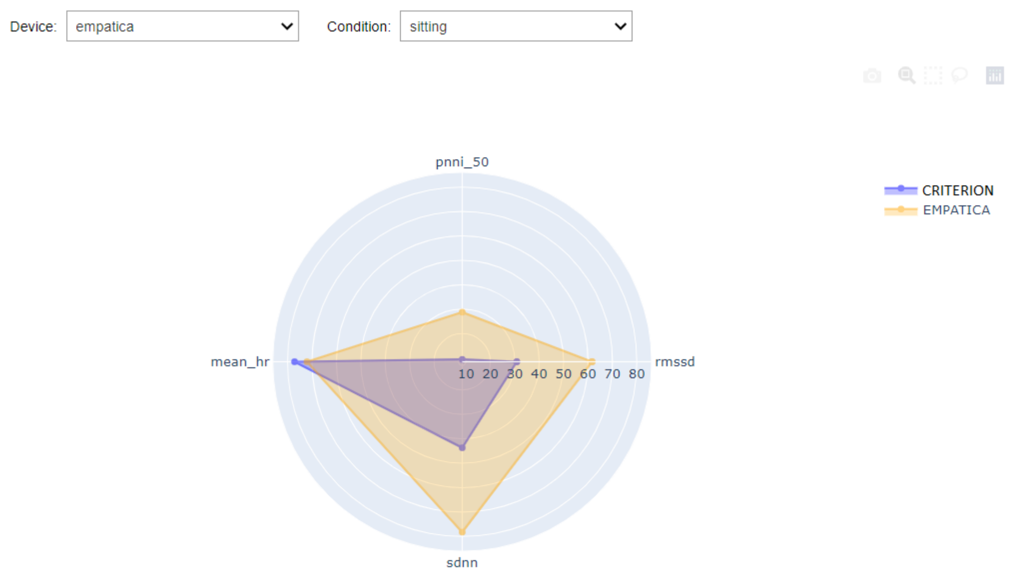

Figure 4: An example of the radar plot at the group level, illustrating a comparison between a criterion device and a specified device for `pnni_50`, `rmssd`, `sdnss` (all distinct extracted features for HRV), and mean heart rate (`mean_hr`). The user can easily switch between devices and conditions by interacting with the `Device` and `Condition` widgets.

Finally, `WearableHRV` allows for detailed comparison between each condition and device against the criterion, using the most commonly used statistical analyses: mean absolute percentage error, regression analysis, ICC, and Bland-Altman analysis through the `mape_analysis`, `regression_analysis`, `icc_analysis`, and `blandaltman_analysis` functions. These analyses are complemented with intuitive plots that establish the validity of the wearables.

Acknowledgements

This study is an output of the ALIVE project (Improving Adolescent mentaL health by reducing the Impact of PoVErty). This work was supported by the Wellcome Trust [221940/Z/20/Z]. We acknowledge contribution from Dr. Denise J. van der Mee and advice from Dr. Marco Altini.

References

Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, *28*(3), R1. https://doi.org/10.1088/0967-3334/28/3/R01

Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: The analysis of method comparison studies. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *32*(3), 307–317. https://doi.org/10.2307/2987937

Bent, B., Goldstein, B. A., Kibbe, W. A., & Dunn, J. P. (2020). Investigating sources of inaccuracy in wearable optical heart rate sensors. *Npj Digital Medicine*, *3*(1), 1–9. https://doi.org/10.1038/s41746-020-0226-6

Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy*, *86*(2), 94–99. https://doi.org/10.1016/S0031-9406(05)61211-4

Challoner, A. V. J., & Ramsay, C. A. (1974). A photoelectric plethysmograph for the measurement of cutaneous blood flow. *Physics in Medicine & Biology*, *19*(3), 317. https://doi.org/10.1088/0031-9155/19/3/003

Gomes, P., Margaritoff, P., & Silva, H. (2019). pyHRV: Development and evaluation of an open-source python toolbox for heart rate variability (HRV). *Proc. Int'l Conf. On Electrical, Electronic and Computing Engineering (IcETRAN)*, 822–828.

Haghayegh, S., Kang, H.-A., Khoshnevis, S., Smolensky, M. H., & Diller, K. R. (2020). A comprehensive guideline for bland–altman and intra class correlation calculations to properly compare two methods of measurement and interpret findings. *Physiological Measurement*, *41*(5), 055012. https://doi.org/10.1088/1361-6579/ab86d6

Hill, L. K., Hu, D. D., Koenig, J., Sollers, J. J., Kapuku, G., Wang, X., Snieder, H., & Thayer, J. F. (2015). Ethnic Differences in Resting Heart Rate Variability: A Systematic Review and Meta-Analysis. *Psychosomatic Medicine*, *77*(1), 16–25. https://doi.org/10.1097/PSY.0000000000000133

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

Makridakis, S. (1993). Accuracy measures: Theoretical and practical concerns. *International Journal of Forecasting*, *9*(4), 527–529. https://doi.org/10.1016/0169-2070(93)90079-3

Nederend, I., Harkel, A. D. J. ten, Blom, N. A., Berntson, G. G., & Geus, E. J. C. de. (2017). Impedance cardiography in healthy children and children with congenital heart disease: Improving stroke volume assessment. *International Journal of Psychophysiology*, *120*, 136–147. https://doi.org/10.1016/j.ijpsycho.2017.07.015

Pinheiro, N., Couceiro, R., Henriques, J., Muehlsteff, J., Quintal, I., Goncalves, L., & Carvalho, P. (2016). Can PPG be used for HRV analysis? *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2016*, 2945–2949. https://doi.org/10.1109/EMBC.2016.7591347

Quintana, D. S., Alvares, G. A., & Heathers, J. a. J. (2016). Guidelines for Reporting Articles on Psychiatry and Heart rate variability (GRAPH): Recommendations to advance research communication. *Translational Psychiatry*, *6*(5), e803–e803. https://doi.org/10.1038/tp.2016.73

Robin Champseix, C. L. C., Laurent Ribiere. (2021). A python package for heart rate variability analysis and signal preprocessing. *Journal of Open Research Software*, *9*, 28. https://doi.org/10.5334/jors.305

Schäfer, A., & Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart rate variability?: A review on studies comparing photoplethysmographic technology with an electrocardiogram. *International Journal of Cardiology*, *166*(1), 15–29. https://doi.org/10.1016/j.ijcard.2012.03.119

Stone, J. D., Ulman, H. K., Tran, K., Thompson, A. G., Halter, M. D., Ramadan, J. H., Stephenson, M., Finomore, V. S., Galster, S. M., Rezai, A. R., & Hagen, J. A. (2021). Assessing the Accuracy of Popular Commercial Technologies That Measure Resting Heart Rate and Heart Rate Variability. *Frontiers in Sports and Active Living*, *3*. https://doi.org/10.3389/fspor.2021.585870