

Reducing Motion Artifacts of Pulse Intervals from Photoplethysmogram of a Commercial Wristband for Heart Rate Variability Analysis*

Hui Wang, Holly Jimison, *Member, IEEE*, and Misha Pavel, *Senior Member, IEEE*

Abstract—Heart rate monitoring based on photoplethysmography (PPG) is a noninvasive and inexpensive way of measuring many important cardiovascular metrics such as heart rate and heart rate variability, and has been used in many wearable devices. Unfortunately, the accuracy of the measurements is compromised by motion artifacts. We propose a theoretically sound method to reduce the motion artifacts of heart rate sensed by a commercial wristband. This method is based on outlier detection and singular spectrum analysis which enables us to reduce the movement-related noise in non-stationary signals. The results suggest that this method exhibits high correspondence to the simultaneously measured heart rate using ECG. Several metrics of heart rate variability computed from cleaned data also indicate high agreement with those obtained from ECG.

I. INTRODUCTION

Heart rate is an important metric reflecting individuals' health and even emotional states, and is therefore a desirable target of continuous monitoring especially using wearable devices. Photoplethysmography (PPG) is an optical measuring technique that measures blood volume changes near the surface of the skin. Many useful cardiovascular metrics can be obtained from the PPG signal including heart rate, blood oxygen saturation, and even estimates of blood pressure[1]. It is also non-invasive and inexpensive compared to clinical methods of measuring the cardiovascular metrics, which leads to the wide adoption of it in wearable devices such as smartwatches and wristbands [2]. Unfortunately, the ambulatory PPG signal is very sensitive to movement-induced noise, which affects the accuracy of the health metrics inferred from the raw signal. Since movements are unavoidable in most practical ambulatory applications, the detection and reduction of motion artifacts in PPG signal is very crucial.

Various methods and techniques have been developed to address this problem [3] [4]. Most of these works focus on obtaining a more accurate estimation of average heart rate [4], instead of instantaneous heart rate (IHR), which is required for estimating Heart Rate Variability (HRV). Heart rate variability is often interpreted in terms of estimates

of the fluctuations between successive heart beats [5] [6], and is considered an important metric for both physical and emotional well-being [7] [8]. Analysis of HRV can be done in time domain, frequency domain, and nonlinear domain. Commonly used time domain HRV metrics include SDNN (standard deviation of normal heart period intervals), RMSSD (root mean square of the differences between successive heart periods) and pNN50 (the percentage of adjacent heart periods that differ by more than 50 ms) [6]. Time domain HRV analysis requires beat by beat heart period data typically obtained from the R peaks of electrocardiogram (ECG). In the case of PPG, beat by beat heart period data can be obtained by detecting systolic peaks of the PPG signal [9].¹ Frequency domain HRV analysis applies under the assumption that heart rate can be considered as a continuous signal, and spectral analysis of this signal can be obtained using Fourier transform or autoregressive modeling. Commonly used frequency domain HRV metrics include low frequency (0.04 Hz - 0.15 Hz) power, high frequency (0.15 Hz - 0.4 Hz) power, and the ratio of low frequency to high frequency power [6]. Average heart rate typically does not preserve the beat-to-beat fluctuations for time domain HRV analysis. It is also an ill-defined variable that lacks theoretical foundation for frequency domain HRV analysis.

In this work, we propose a method for reducing motion artifacts of pulse intervals obtained from the PPG of a commercial wristband within a theoretical framework of IHR and HRV. We evaluate the IHR obtained from PPG against that obtained from ECG, and showed Pearson correlation is much higher and the Average Absolute Error (AAE) is much lower after cleaning. Some of the HRV metrics computed from the cleaned IHR also showed good agreement with that obtained from ECG. This paper is organized as follows. In section II we describe the theoretical framework. In section III we describe our approach in detail. In section IV we present the results of our cleaning method and the evaluation of these results. Lastly, we discuss the limitations of this work and future directions in section V.

II. THEORETICAL FRAMEWORK

The integral pulse frequency modulation (IPFM) model was used by many researchers to model the regulation of heart rate by the autonomic nervous system (ANS) [10] [11]

¹Although the lag between the R peak of ECG to the systolic peak of PPG of the same cardiac cycle may vary and contributes to the difference between HRV computed from ECG and PPG, this variance is likely to be negligible compared to the variance caused by motion artifacts. In this work, we assume this lag to be approximately constant.

*This research was funded by the National Institute on Disability, Independent Living and Rehabilitation Research for the LiveWell Rehabilitation Engineering Research Center (Grant 90RE5028) and National Science Foundation Smart and Connected Health: Self-powered Smart Ring for Always-On Health Interventions (Grant 2014556).

H. Wang is with the Consortium on Technology for Proactive Care (CTPC), and the ARTificially Intelligent Social Agents Lab (ARTISAL) at Northeastern University, Boston, MA 02115 USA (phone: 412-721-4248; email: huiwang@ccs.neu.edu)

H. Jimison and M. Pavel are with the CTPC, Northeastern University, Boston, MA 02115 USA (emails: h.jimison@northeastern.edu and m.pavel@northeastern.edu)

[12]. The IPFM model takes a continuous signal as input, and generates a spike when the integral of the input signal reaches a threshold. The integral is reset each time the threshold is reached. Therefore, it transforms a continuous signal to a sequence of events.

We define IHR as a continuous function $v(t)$. Cardiac cycles can be represented by a sequence of events typically considered to be the peaks of R-wave in ECG, or the systolic peaks in PPG. This event sequence R_i , ($i = 1, 2, \dots, N$), is generated by an IPFM model taking $v(t)$ as the input signal. The i -th event R_i is generated at time T_i when the integral reaches a fixed threshold θ , as described in (1).

$$\int_{T_{i-1}}^{T_i} v(t) dt = \theta \quad (1)$$

Under this assumption we define the time between two consecutive events R_{i-1} and R_i as pulse interval RR_i , represented in (2).

$$RR_i = T_i - T_{i-1} \quad (2)$$

The heart rate variability in this model is captured by the function $v(t)$. The variation of $v(t)$ can be attributed to the regulation of the ANS through the sinoatrial node.

Typically, $v(t)$ is not directly measurable, but it can be inferred indirectly through the measurable event sequences R_i it generates. Based on the negative relationship between samples of $v(t)$ and RR_i (higher values of $v(t)$ correspond to shorter RR_i), the reciprocals of RR_i can be used to approximate $v(t)$. RR_i sequence is usually upsampled to obtain a smooth approximation. Suppose V_i is a sequence of evenly spaced samples of $v(t)$. V_i can be estimated through (3), where SS_i sequence is the upsampled RR_i sequence, f is the resampling frequency, and N is the length of RR_i sequence.

$$V_i = \frac{1}{SS_i}, (i = 1, 2, \dots, fN). \quad (3)$$

III. METHOD

A. Overview

The input of our method is raw RR_i sequence containing outliers and artifacts, as well as accelerometer data recorded simultaneously, and the output is cleaned V_i sequence. The pipeline of this method is shown in Figure 1. First, outliers among the RR_i sequence are detected using rules recommended in [13]. We assume small outliers among the RR_i sequence are caused by erroneously detected extra beats, and large outliers are caused by missing beats failed to be detected. After outlier detection, we remove the small outliers and impute the large outliers. The RR_i sequence is then transformed to the V_i sequence through upsampling and reciprocal transformation. Finally, Singular Spectrum Analysis is applied to further smooth the segment when movement intensity is high according to the accelerometer data.

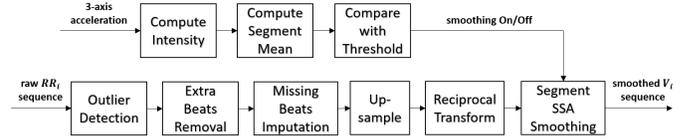


Fig. 1. Pipeline of Method

B. Outlier Detection

We implemented a set of rule-based outlier detection criterion originally designed for ECG in [13]. The rules are based on the distribution of the difference between two successive RR_i . A criterion of interval difference is computed based on quartile deviation. Successive interval differences that exceed this criterion are considered to be caused by an outlier. We found this rule-based criterion to be reasonably effective on noisy RR_i sequence from PPG. The only modification we made is merging successive outliers into one when multiple successive RR_i are identified to be outliers. These segments of multiple outliers are usually caused by sudden movement which tends to corrupt the PPG signal and makes peak detection difficult.

C. Extra Beats Removal

We assume small outliers among the RR_i sequence are caused by erroneously detected extra beats. Small outliers are identified from all the detected outliers using robust mean estimated by Minimum Covariance Determinant (MCD) [14]. We consider a detected outlier to be a small outlier for removal if the nearest integer of the outlier RR divided by robust mean is zero. The small outlier is removed by adding its value to the smaller one of the two neighboring RR_i .

D. Missing Beats Imputation

We assume large outliers are caused by missing beats failed to be detected. We identify large outliers from all the outliers using robust mean computed the same way as small outliers. We consider a detected outlier for as a large outlier for imputation if it satisfies (4). The large outlier is then split into smaller pulse intervals using an autoregressive (AR) model. The outliers that are larger than $\overline{RR_i}$ with the ratio between it and $\overline{RR_i}$ rounded to one are left untreated at this step since splitting them would creating small outliers. For the large outlier ready for imputation, the number of missing pulse intervals is estimated by the left side of (4). An AR model is fitted to a window of samples centered around the large outlier (AR order = 4, sample size = 40). Each missing pulse interval is estimated sequentially using the previous intervals and the obtained AR coefficients. After all the missing intervals are estimated by the AR model, a weight is multiplied to them to make sure the total length of them is equal to the value of the large outlier.

$$\text{round}\left(\frac{RR_{\text{outlier}}}{\overline{RR_i}}\right) \geq 2 \quad (4)$$

E. Smoothing using Singular Spectrum Analysis

After removing extra beats and imputing missing beats, the RR_i sequence no longer has extreme outliers, but still has artifacts, especially in segments where the magnitude of acceleration is large. We apply Singular Spectrum Analysis (SSA) to these segments in order to further reduce these artifacts. SSA is essentially a Principal Component Analysis on time series [15]. The aim of SSA is to decompose a time series into the sum of a small number of interpretable components such as trend, periodicities and noise [16]. There are two stages in SSA, decomposition and reconstruction. Given a discrete time series $x[n]$, $n = 1, 2, \dots, M$, a trajectory matrix X of dimension $L \times K$ is created from K lagged windows of length L segments of the original signal (L and K satisfy $2 \leq L \leq N/2$ and $K = M - L + 1$ respectively). In the decomposition stage, the trajectory matrix X is converted into the sum of d elementary matrices X_i with rank 1 via singular value decomposition (SVD), as described in (5), where d is the rank of the trajectory matrix, σ_i is the i th singular value, U_i and V_i in this equation are the i th column of matrices U and V from SVD. In the reconstruction stage, elementary matrices X_i are grouped into components that can be considered to roughly represent trend, periodicities and noises. Typically the elementary matrix with the largest singular value corresponds to the trend in the time series, while the ones with small singular values usually correspond to noise. The noise component can be excluded from the summation to obtain a smoother reconstructed signal.

$$X = U\Sigma V^T = \sum_{i=0}^{d-1} \sigma_i U_i V_i^T = \sum_{i=0}^{d-1} X_i \quad (5)$$

SSA typically applies to evenly sampled time series, and thus cannot be performed on pulse intervals directly. Instead, we apply SSA on an evenly sampled IHR signal. Based on our theoretical framework in section II, IHR is defined as a continuous signal $v(t)$ representing the regulation of ANS on heart rate. We use the reciprocals of interpolated RR_i to approximate $v(t)$. The RR_i sequence is upsampled to 4 Hz, and the V_i sequence is obtained using (3).

We apply SSA to non-overlapping segments of V_i . We choose the number of elementary matrices used for reconstruction based on the acceleration signal recorded at the same time. Our previous work has shown that the higher the acceleration, the lower the agreement between the IHRs obtained from simultaneously recorded PPG and ECG signal [17]. We define movement intensity as $\left| \sqrt{a_x^2 + a_y^2 + a_z^2} - 1 \right|$, where a_x, a_y, a_z are acceleration in each axis in unit g. When the mean movement intensity in a segment is above a threshold, we retain the first elementary matrix X_0 for reconstruction. When the mean movement intensity is below the threshold, we consider no motion artifact is present in this segment and skip the smoothing in order to retain as much inherent variance in the pulse intervals as possible. This threshold is chosen empirically to separate out segments severely affected by motion. The

length of each segment was chosen to be 10 seconds in consideration of a balance between capturing the change of acceleration and having enough data for SSA. The window length L for SSA is set to be 1/3 of the number of points in the segment.

IV. RESULTS

A. Dataset

Before discussing results, we briefly describe the study that collected the dataset being used for testing our method. 9 healthy participants went through interleaving stress and relaxation tasks during a lab session which lasted about an hour. Details of the study protocol can be found in [17]. The study was approved by the Institutional Review Board (IRB) of Northeastern University. The approach we proposed is suitable for reducing motion artifacts caused by mild and sporadic motions, which were present in the data of all the subjects with varying degrees. Throughout the study, the participants were wearing an ECG recording device (Firstbeat Bodyguard 2) and two wrist bands with reflective PPG sensors (Microsoft Band 2 and Empatica E4). This work uses the pulse intervals data and accelerometer data from Microsoft Band, and evaluate the results against the heart beat intervals data from Firstbeat.

B. Evaluation of Instantaneous Heart Rate

Beat-by-beat comparison directly between the RR_i sequences from ECG and PPG is not feasible here, because one-to-one mapping cannot be easily found due to the outliers and artifacts in the RR_i sequence from PPG. Instead, we compare the samples of the IHR signal, V_i obtained using (3) where f is set to 4 Hz. The unit of IHR is beats per second when the unit of RR_i is seconds. We converted it to a more commonly used unit beats per minutes (bpm) by multiplying 60 to it. The V_i sequences from PPG and that from ECG were aligned by finding the lag between them using cross correlations. We then compute Pearson Correlation (Corr) and Average Absolute Error (AAE) between the two aligned V_i sequences. The cleaning method was applied to the first set of relaxation and stress tasks (listening to relaxing music, viewing and describing neutral and evocative images). The results are shown in Table I. The raw and cleaned signals of a representative subject is shown in Figure 2.

It can be seen from Table I that AR imputation performs slight worse compared to an simple equal split imputation approach. This remains to be true after different orders (2,4,6,8) and window size (20, 40, 60, 80) of the AR model were tested. After inspection of the plots, we found that for most of the long outliers, AR imputation failed to recover the true value of the missing pulse intervals, and therefore creates more variance than the equal split imputation.

SSA improves the results of all the subjects. Before applying SSA, RR_i sequence is upsampled to 4 Hz. We experimented with linear interpolation and cubic interpolation, and found linear interpolation performs better. We also tested the ordering between reciprocal transform and SSA, and found that applying reciprocal transform before SSA leads to better

TABLE I
CORRELATION AND AAE BETWEEN IHR FROM PPG AND ECG

ID	Raw		AR		EqualSplit		EqualSplit+SSA	
	Corr	AAE	Corr	AAE	Corr	AAE	Corr	AAE
1	0.247	6.351	0.436	5.389	0.439	5.353	0.482	4.970
2	0.323	7.241	0.593	5.723	0.608	5.587	0.642	5.264
3	0.178	7.632	0.350	5.942	0.351	5.890	0.384	5.505
4	0.681	3.344	0.762	3.067	0.769	3.031	0.791	2.933
5	0.282	5.569	0.477	4.180	0.491	4.109	0.532	3.689
6	0.222	6.498	0.503	4.891	0.515	4.810	0.545	4.498
7	0.172	3.202	0.366	2.524	0.385	2.452	0.486	1.756
8	0.442	3.286	0.598	2.518	0.606	2.511	0.649	2.245
9	0.005	6.029	0.469	3.138	0.481	3.054	0.497	2.856

¹ Corr: Pearson's correlation coefficient ² AAE: Average Absolute Error in bpm

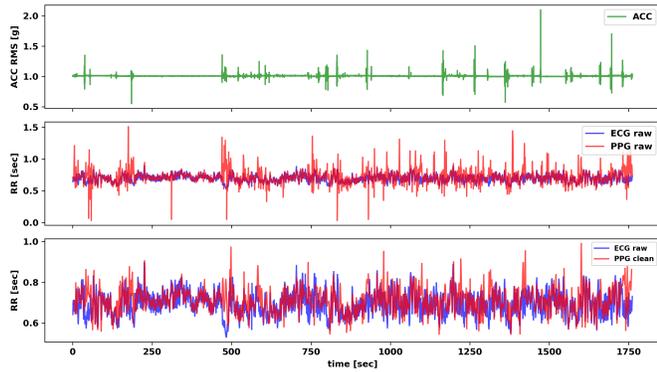


Fig. 2. A segment of raw and cleaned pulse intervals with acceleration magnitude from a representative subject

results that applying it after SSA. During SSA reconstruction stage, we use only the first component for reconstruction based on our observation that when movement intensity is high, the main difference between the RR_i sequences from PPG and ECG are in the trend, and using more components for reconstruction does not help reducing this difference.

C. Evaluation of Heart Rate Variability

In this section, we compare commonly used time domain and frequency domain HRV metrics obtained from PPG after cleaning with that from ECG. For each subject, we use a 5-minutes segment centered at the middle point of the viewing evocative pictures task, which is the only task that lasts longer than 5 minutes other than the relaxation task. Five minutes is the standard duration for short-term HRV analysis [6]. The evocative picture viewing task also has more motion artifacts present in the raw signals compared to the relaxation task, which makes it more suitable to demonstrate our cleaning method.

Time domain HRV analysis requires the RR_i sequence. The output of our cleaning method is an evenly sampled IHR signal, which cannot be used directly to compute the time domain HRV metrics. The IPFM model can be used to simulate RR_i from IHR. We upsample the IHR signal to 1000 Hz in order to obtain 1 millisecond resolution of the generated pulse intervals. Although the generated RR_i

sequences do not match the original RR_i sequences exactly, they are very close to each other as shown by a segment of a representative subject in Figure 3. Since this part is not the focus of this work, the results of the comparison between original RR_i sequences and generated RR_i sequences are not reported here for the sake of space.

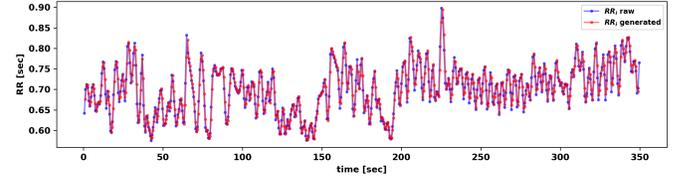


Fig. 3. Original and generated RRs from ECG

The Pearson's correlation between the HRV metrics computed from ECG RR_i and those from generated PPG RR_i after cleaning are reported in Table II. SDNN, pNN50 and low frequency power show significant correlation ($p < 0.05$) between the ECG and PPG metrics.

Bland-Altman agreement analysis [18] of the HRV metrics are shown in Figure 4. RMSSD, pNN50, and HF power have good agreement. SDNN and LF power have a negative bias, but the bias is fairly small considering the range of the metric. The biases of HF n.u. and LF/HF are likely to be aggravated by the division operation. Most subjects are within the 95 percentile ($+1.96SD$) of the differences. The outlier exceeding 95 percentile in RMSSD, pNN50 and HF plots actually represents the same subject (Subject 2). Plots of acceleration data revealed that Subject 2 exhibited constant movements during the task, which differ considerably from the sporadic movements exhibited by other subjects.

TABLE II
CORRELATION BETWEEN HRVs FROM ECG AND CLEANED PPG

HRV	Mean \pm Std (ECG)	Mean \pm Std (PPG)	Corr	p
meanRR	801.51 \pm 130.64	800.80 \pm 122.53	0.99	0.00
SDNN	51.68 \pm 21.22	57.23 \pm 23.75	0.68	0.04
RMSSD	717.69 \pm 371.66	754.95 \pm 310.87	0.50	0.16
pNN50	0.17 \pm 0.12	0.16 \pm 0.13	0.71	0.03
LF power	854.48 \pm 620.94	911.97 \pm 522.01	0.68	0.04
HF power	503.22 \pm 612.72	533.82 \pm 378.54	0.51	0.16
HF n.u.	0.30 \pm 0.13	0.36 \pm 0.10	0.65	0.06
LF/HF	2.99 \pm 1.87	2.13 \pm 1.39	0.30	0.44

¹ The unit of meanRR, SDNN and RMSSD are *ms*; the unit of LF power and HF power are *bpm*²; pNN50, LF/HF and HF n.u. are ratios. ² HF n.u. is high frequency power in normalized unit, computed as HF/(LF+HF). Since LF n.u. is equivalent to 1 - HF n.u., it is not reported.

V. DISCUSSION

In this work, we proposed a method for reducing motion artifacts of pulse intervals obtained from PPG. We first detected outlier beats, removed extra beats and imputed missing beats. Then SSA smoothing was applied to segments with high movement intensity based on accelerometer data. The agreement between the cleaned IHR from PPG and that from ECG is greatly improved for all the subjects. We also

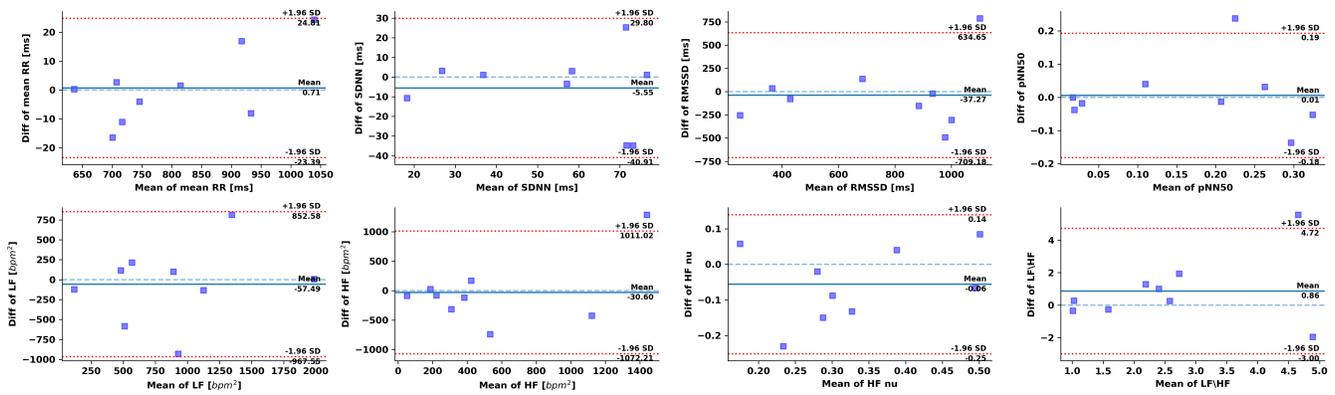


Fig. 4. Bland-Altman agreement plots between HRV metrics from ECG and cleaned PPG

evaluated HRV metrics, and showed SDNN, pNN50 and LF power significantly correlated with those obtained from ECG. In addition to the good performance, our approach has an advantage of the underlying theoretical framework that defines IHR and HRV mathematically, and allows the reconstruction of RR_i sequence from IHR. Overall, this paper shows the potential of using commercial wearables with PPG sensor for HRV analysis under mild motions, and is an important step towards continuous monitoring of emotional well-being in ambulatory settings.

There are several limitations in this work. First, our method starts with low quality pulse intervals because the raw PPG signal from Microsoft Band 2 was not available. The raw PPG signal contains more information and therefore has more room for improvement in reducing motion artifacts. Second, we used a fixed threshold computed from accelerometer data for localized SSA smoothing. Multiple thresholds, a regression model or probabilistic approach may be more effective. Going forward, we intend to address these limitations. We plan to extend this method to include a function of getting pulse intervals from raw PPG signal and test it on raw the PPG from a research device (Empatica E4). We also plan to assess and improve this method using data collected during diverse conditions such as physical activities and mental stress.

ACKNOWLEDGMENT

We would like to thank Xuan Li, Maciej Kos, Christine Gordon and Iman Khaghani-Far for conducting the study to collect the data set we use, as well as maintaining and sharing it.

REFERENCES

- [1] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," *Physiological measurement*, vol. 28, no. 3, p. R1, 2007.
- [2] T. Tamura, Y. Maeda, M. Sekine, and M. Yoshida, "Wearable photoplethysmographic sensors past and present," *Electronics*, vol. 3, no. 2, pp. 282–302, 2014.
- [3] D. Pollreisz and N. TaheriNejad, "Detection and removal of motion artifacts in ppg signals," *Mobile Networks and Applications*, pp. 1–11, 2019.

- [4] D. Biswas, N. Simões-Capela, C. Van Hoof, and N. Van Helleputte, "Heart rate estimation from wrist-worn photoplethysmography: A review," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6560–6570, 2019.
- [5] U. R. Acharya, K. P. Joseph, N. Kannathal, C. M. Lim, and J. S. Suri, "Heart rate variability: a review," *Medical and biological engineering and computing*, vol. 44, no. 12, pp. 1031–1051, 2006.
- [6] F. Shaffer and J. Ginsberg, "An overview of heart rate variability metrics and norms," *Frontiers in public health*, vol. 5, p. 258, 2017.
- [7] J. F. Thayer, S. S. Yamamoto, and J. F. Brosschot, "The relationship of autonomic imbalance, heart rate variability and cardiovascular disease risk factors," *International journal of cardiology*, vol. 141, no. 2, pp. 122–131, 2010.
- [8] A. H. Kemp and D. S. Quintana, "The relationship between mental and physical health: insights from the study of heart rate variability," *International Journal of Psychophysiology*, vol. 89, no. 3, pp. 288–296, 2013.
- [9] A. Schäfer and J. Vagedes, "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*, vol. 166, no. 1, pp. 15–29, 2013.
- [10] J. Mateo and P. Laguna, "New heart rate variability time-domain signal construction from the beat occurrence time and the ipfm model," in *Computers in Cardiology 1996*, pp. 185–188, IEEE, 1996.
- [11] R. D. Berger, S. Akselrod, D. Gordon, and R. J. Cohen, "An efficient algorithm for spectral analysis of heart rate variability," *IEEE Transactions on biomedical engineering*, no. 9, pp. 900–904, 1986.
- [12] R. W. Deboer, J. M. Karemaker, and J. Strackee, "Comparing spectra of a series of point events particularly for heart rate variability data," *IEEE Transactions on Biomedical Engineering*, no. 4, pp. 384–387, 1984.
- [13] G. G. Berntson, K. S. Quigley, J. F. Jang, and S. T. Boysen, "An approach to artifact identification: Application to heart period data," *Psychophysiology*, vol. 27, no. 5, pp. 586–598, 1990.
- [14] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
- [15] D. H. Schoellhamer, "Singular spectrum analysis for time series with missing data," *Geophysical research letters*, vol. 28, no. 16, pp. 3187–3190, 2001.
- [16] H. Hassani, "Singular spectrum analysis: methodology and comparison," 2007.
- [17] M. Kos, X. Li, I. Khaghani-Far, C. M. Gordon, M. Pavel, and H. B. Jimison, "Can accelerometry data improve estimates of heart rate variability from wrist pulse ppg sensors?," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1587–1590, IEEE, 2017.
- [18] D. Giavarina, "Understanding bland altman analysis," *Biochemia medica*, vol. 25, no. 2, pp. 141–151, 2015.