End-to-End PPG Processing Pipeline for Wearables: From Quality Assessment and Motion Artifacts Removal to HR/HRV Feature Extraction

Mohammad Feli[1], Kianoosh Kazemi[1], Iman Azimi[2], Yuning Wang[1], Amir M. Rahmani[2,3], and Pasi Liljeberg[1]
[1]*Department of Computing, University of Turku, Turku, Finland*
[2]*Department of Computer Science, University of California, Irvine, USA*
[3]*School of Nursing, University of California, Irvine, USA*
{mohammad.feli, kianoosh.k.kazemi, yuning.y.wang, pasi.liljeberg}@utu.fi, {azimii, a.rahmani}@uci.edu

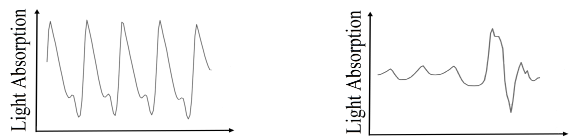
***Abstract*—The rapid development of wearable technology has enabled remote photoplethysmography (PPG)-based health monitoring in everyday settings, offering real-time and continuous monitoring of cardiovascular parameters, such as heart rate (HR) and heart rate variability (HRV). However, PPG signals collected in daily life are prone to artifacts and noise, posing challenges to HR and HRV extraction. The existing HR and HRV extraction methods cannot effectively handle noisy PPG signals and ensure accurate results. Additionally, current Python packages were primarily designed for analyzing "clean" PPG signals, limiting their performance in handling artifacts and noise and resulting in unreliable HR and HRV measurements. In this paper, we propose a robust end-to-end PPG processing pipeline to reliably extract HR and HRV from PPG signals collected in free-living settings. The pipeline comprises three machine learning-based PPG analysis methods: signal quality assessment, reconstruction of noisy signal, and systolic peak detection. We assess the proposed PPG pipeline using a dataset including PPG and Electrocardiogram (ECG) signals recorded from 46 individuals by smartwatches. Our evaluation demonstrates the proposed pipeline's superior performance compared to two established benchmark methods in terms of correlation and mean absolute error with ECG as the reference. We also provide the Python implementation of our pipeline for the research community to facilitate integration into their solutions.**

***Index Terms*—Photoplethysmography, Heart rate, Heart rate variability, Wearable devices, Health monitoring**

I. Introduction

Thanks to the advances in wearable technology, remote photoplethysmography (PPG)-based health monitoring has been proliferating over the past few decades. PPG is an optical method to measure blood volume changes in the microvascular bed of tissue. PPG signal is recorded by emitting light to the skin and capturing the reflected or transmitted light using photodetectors [1]. The method is easy to implement and non-invasive, widely used in smartwatches and fitness trackers to monitor the cardiovascular system. PPG can provide real-time and continuous monitoring of various physiological parameters, such as heart rate (HR) and heart rate variability (HRV),

This research was supported by the Academy of Finland through the SLIM Project (grant numbers 316810 and 316811) and the U.S. National Science Foundation through the UNITE Project (grant number SCC CNS-1831918).



(a) Clean PPG (b) Noisy PPG

Fig. 1: PPG Samples. (a) Signal is clean. (b) Signal is noisy.

enabling healthcare professionals to track changes and patterns over time [2]. Fig. 1a shows a view of a PPG signal.

While the PPG enables remote vital signs monitoring, its signal quality is significantly affected by artifacts and noise, particularly when collected in free-living conditions. In such scenarios, PPG signals are vulnerable to corruption due to various artifacts that arise when subjects engage in different physical activities (see Fig. 1b). Moreover, environmental factors (e.g., ambient light) can significantly distort PPG signals. Given the inevitable presence of such noises in real-world applications, extracting vital signs, such as HR and HRV, from PPG becomes a significant challenge.

Numerous studies have been conducted to extract HR and HRV from PPG signals [3]. Conventional PPG analysis methods leverage signal processing techniques. The TROIKA method [4] was proposed, as one of the leading traditional approaches, to estimate HR by employing a spectral matrix constructed from PPG and acceleration signals and a spectral peak tracking method. Inspired by TROIKA, JOSS [5] was introduced, demonstrating superior performance for HR estimation. Signal decomposition techniques were also used [6] to obtain HR and HRV from PPG.

Moreover, machine learning (ML) methods have been employed for PPG-based HR extraction. Several ML approaches have been proposed and evaluated via the TROIKA dataset [4], including data from specific physical activities. For example, a Random Forest method [7] was proposed to estimate the position of systolic peaks in PPG signals. Another approach [8] was introduced to track HR and HRV during physical activities using a Multi-layer Perceptron Neural Network. Recently, deep learning techniques [9] have been explored for HR and HRV extraction from PPG. For instance, Kazemi et al al. [10] have proposed a convolutional neural network (CNN) method to detect systolic peaks in PPG data from free-living conditions, which were then used for HR and HRV estimation.

In addition to the existing studies, several Python packages offer tools for the analysis of PPG signals and the extraction of HR and HRV. For instance, Neurokit [11] and BioSPPy [12] provide various processing functions, such as PPG signal peak detection. Similarly, HeartPY [13] was developed for PPG signal analysis, offering visualization, preprocessing, filtering, and peak detection methods.

Many existing PPG-based methods for HR and HRV extraction ignore the impact of noise and motion artifacts in their analyses [3]. Consequently, they cannot consistently guarantee accurate results, especially in cases with a low Signal-to-Noise ratio. Similarly, the available Python packages [11, 12, 13] were primarily designed for analyzing "clean" PPG signals, making them ineffective in handling motion artifacts and noise. Some studies attempt to address this issue by utilizing accelerometer signals to capture motion artifacts [4, 5, 6] and predefined activities [7, 8, 9]. However, they struggle with unseen noisy signals that significantly distort PPG waveforms, as these methods lack a robust strategy to handle noisy PPG signals effectively. A few studies have developed methods for free-living PPG data [10], but they merely focused on peak detection, neglecting artifacts and noise handling.

Recently, numerous studies have harnessed PPG-enabled wearable devices to extract both short-term and long-term HR and HRV data, serving diverse purposes including stress assessment [14], sleep quality evaluation [15], mental health monitoring [16], and maternal health assessment [17]. Given the critical nature of these health applications, ensuring the acquisition of accurate and dependable HR and HRV is crucial. Shortcomings in performance could potentially compromise patient safety, decision-making, and treatment effectiveness. Our objective is to tackle this problem by developing a noise-resilient and robust methodology tailored for these applications, enabling the extraction of HR and HRV from PPG signals reliably, even in the presence of noise.

In this paper, we propose an end-to-end pipeline to extract HR and HRV from PPG signals collected in free-living conditions. The pipeline consists of three major ML-based PPG analysis methods. Firstly, it discriminates between clean and noisy PPG signals, considering the signal's morphology. Secondly, it reconstructs the noisy parts by capitalizing on preceding patterns in the signal. Lastly, it employs a CNN-based approach to perform systolic peak detection and HR and HRV extraction. We assess the proposed PPG pipeline using a dataset including PPG and Electrocardiogram (ECG) signals from 46 individuals recorded during their daily routines. We compare the proposed PPG pipeline with two established benchmark methods: Neurokit [11] and HeartPY [13]. Leveraging ECG as the reference, we assess the methods in terms of correlation and error of HR and HRV parameters. Additionally, we make the pipeline available in Python on GitHub[1] for integration into the research community's solutions.

[1] https://github.com/HealthSciTech/E2E-PPG

II. Dataset

We utilize a dataset collected from a remote health monitoring study [18]. The study recruitment was carried out from July to August 2019 in southern Finland. Forty-six participants (i.e., 23 females and 23 males) were selected for the study. Eligibility criteria included ages 18 to 55 and no cardiovascular disease. They were instructed to wear a Samsung Gear Sport smartwatch [19] on their non-dominant hand and a Shimmer3 ECG chest strap [20] continuously for one day to record PPG and ECG signals while engaging in their daily activities.

Shimmer3 ECG [20] is a portable device that continuously records 12-channel ECG signals for up to 24 hours. ECG signals were collected 24 hours through four limb electrodes positioned on the left arm, right arm, left leg, and right leg. In this study, we only use Lead II ECG for the analysis. Samsung Gear Sport watch [19] is a lightweight and waterproof smartwatch equipped with optical and inertial measurement unit sensors, running on the open-source Tizen operating system. We programmed the watch to record 16-minute PPG signals every 30 minutes. Note that we collected ECG and PPG at sampling frequencies of 512 Hz and 20 Hz, respectively.

Ethics: This research followed ethical guidelines as per the Declaration of Helsinki and the Finnish Medical Research Act (#488/1999). The study received approval from the University of Turku's Ethics Committee for Human Sciences (Statement #44/2019). Participants were informed and consented voluntarily, with the right to withdraw without explanation.

III. Proposed PPG Pipeline

In this section, we propose a novel PPG pipeline approach to extract reliable HR and HRV from PPG. Our objective is to develop an end-to-end pipeline method to effectively handle motion artifacts and noise and extract accurate HR and HRV measurements. The proposed PPG pipeline is illustrated in Fig. 2. First, we implement a preprocessing stage to filter the input (raw) PPG signal and discard frequencies outside the heartbeat frequency range. Subsequently, a PPG signal quality assessment (SQA) model evaluates the signal quality, generating an array of confidence values for each sample in the signal. Any noisy part with a duration less than a specific threshold is then reconstructed using a PPG reconstruction model, followed by reapplying the SQA model to ensure that the reconstructed signal is not distorted. Next, a PPG peak detection model is employed, and interbeat intervals (IBIs) are identified. Finally, HR and HRV parameters are extracted and outputted as vital signs of interest. We elaborate on the various components of the proposed PPG pipeline in the following.

Preprocessing: This stage involves filtering the input raw PPG signal to remove undesired frequencies. PPG signals in their raw form often contain interference components, such as baseline wander and high-frequency noises that do not represent the features of interest (i.e., HR and HRV). To discard these components, we apply a second-order Butterworth

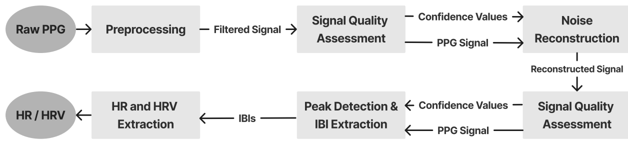
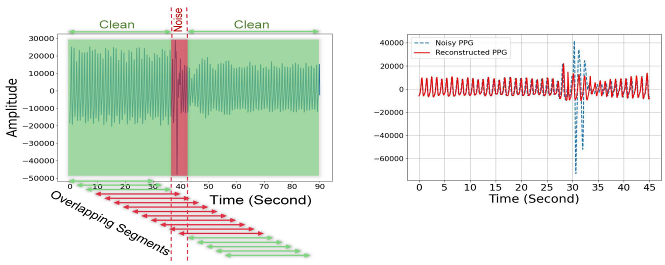


Fig. 2: Proposed PPG pipeline



(a) (b)

Fig. 3: (a) Moving window segmentation followed by SQA. Reliable segments (green) are aggregated to identify clean parts. The remaining gap indicates the noise within the signal. (b) A noisy PPG (blue) and its reconstructed signal (red).

bandpass filter which allows frequencies within the range of 0.5 to 3 Hz to pass but rejects frequencies outside.

Signal Quality Assessment: In this step, we develop a Signal Quality Assessment (SQA) method to identify clean and noisy parts within the PPG signals. Due to the frequent distortion of PPG signals by motion artifacts and noise, it is essential to accurately assess the signal quality before extracting vital signs. Our SQA approach requires PPG signals in a fixed length, which necessitates segmenting the input signals. To this end, we develop a moving window segmentation technique, where the PPG signals are divided into overlapping segments, each spanning 30 seconds, by sliding a window over the signal with a 2-second shifting step.

We evaluate the quality of PPG segments using our previously well-established SQA approach presented in [21]. This method consists of two main phases: PPG feature extraction and classification. Five features are extracted from the PPG segments, including the interquartile range, the standard deviation of power spectral density, the range of energy of heart cycles, and two template matching features as the average Euclidean distances between a template and heart cycles, and correlations between a template and heart cycles. The template is an average of all heart cycles within a given segment. A one-class support vector machine model is then employed to classify the PPG segments into “Reliable” (clean) and “Unreliable” (noisy) classes.

By aggregating the “Reliable” segments, we effectively identify the clean parts within the PPG signal. The resulting gaps between clean parts indicate the presence of noisy parts (see Fig. 3a). The SQA method generates a confidence values array, indicating the quality of each sample in the signal.

Noise Reconstruction: We reconstruct noisy parts within PPG signals that are less than a specific threshold (i.e., 15 seconds in our setup). PPG is quasi-periodic, as it shows the rhythmic activity of the cardiovascular system. When a small noise occurs in the signal, the corrupted part can be reconstructed by exploiting the information in the preceding clean part. We employ our deep convolutional generative adversarial network (GAN) for PPG reconstruction [22].

The deep convolutional GAN consists of a generator and a discriminator. During training, the generator learns patterns and features in clean PPG signals and produces succeeding clean signals. The discriminator then acts as a binary classifier to distinguish the genuine and generated signals. This adversarial training process ultimately leads to the joint optimization of the model performance by minimizing the difference between the reconstructed and original ones.

The trained generator of the model is used to reconstruct the corrupted part of the PPG signal. A sliding window with a fixed length was implemented to feed the PPG signals to the generator. In each iteration, the generator was fed a window of the PPG signal to estimate its clean succeeding points. The iterations were repeated, wherein the sliding window was shifted until the entire corrupted signal was covered.

In our setup, the size of the input window is 15 seconds, the size of the estimated succeeding points is 5 seconds, and the shift size is 2.5 seconds. This process can be iterated to reconstruct the noise for up to 15 seconds. Fig. 3b demonstrates a noisy PPG signal with its reconstructed signal.

Peak Detection and IBI Extraction: We identify systolic peaks in PPG signals, enabling the extraction of IBI values that serve as the basis for obtaining HR and HRV. IBI represents the time duration between two consecutive heartbeats and is computed by measuring the time interval between systolic peaks within the PPG signals.

We perform PPG peak detection using a deep-learning-based method that we presented in [10]. The method has shown superior performance when dealing with noisy signals. The method’s success is attributed to its dilated Convolutional Neural Networks (CNN) architecture, which is capable of processing time-series data efficiently due to the large receptive field provided by the dilated convolutions.

The model takes the PPG signal as an input and produces a probability output that indicates the likelihood of a signal point being a systolic peak. A peak finder function is then applied to identify the peaks’ locations in the signal. The peak finder function first generates a list of all points in the signal that have a probability value above a pre-defined threshold determined experimentally. Subsequently, the function uses a local maximum finder to identify the peaks’ locations. To ensure accuracy, we adopt a rule-based filter, wherein a peak is detected as a false peak if the variation of its NN interval exceeds 20% of the average NN intervals. Subsequently, if the number of false peaks is more than 30% of the total number of systolic peaks in a given segment, the entire segment is considered unreliable and discarded.

HR and HRV extraction: HR and HRV parameters are computed from the IBI values derived in the previous step. As

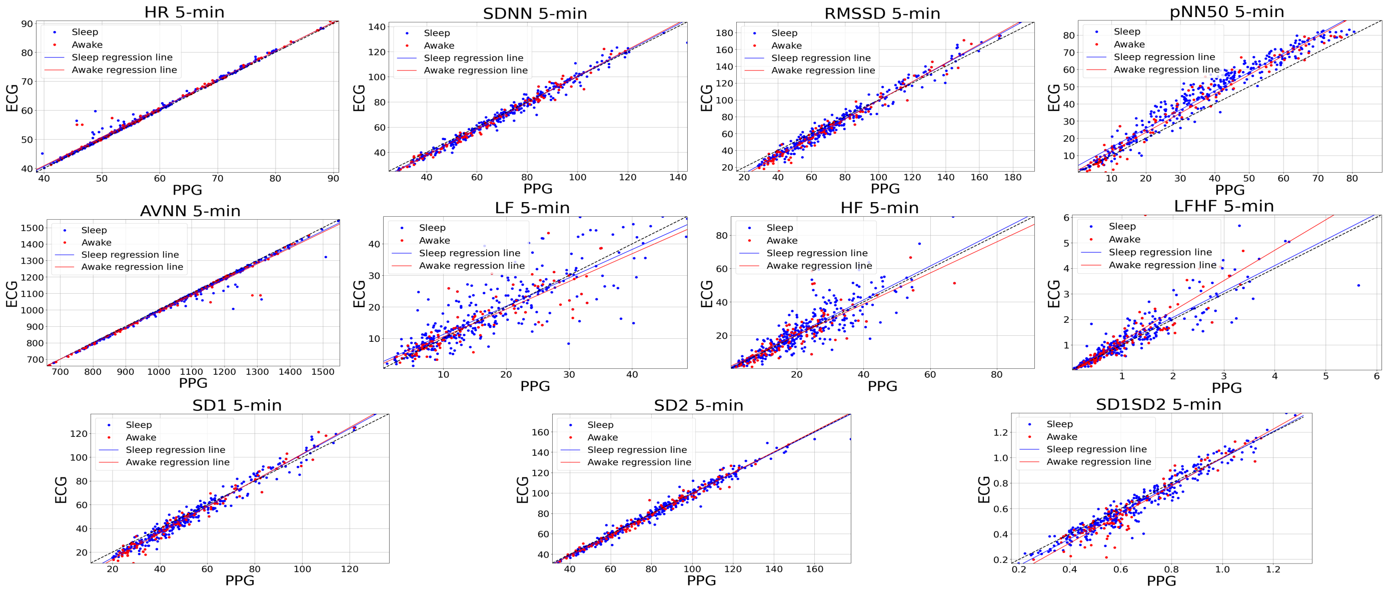


Fig. 4: Regression analysis of 5-min HR and HRV derived by proposed PPG pipeline and reference ECG

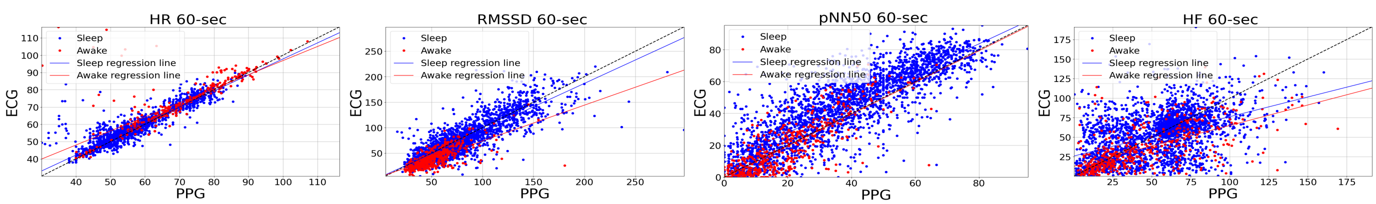


Fig. 5: Regression analysis of 60-sec HR and HRV derived by proposed PPG pipeline and reference ECG

discussed in [23], the parameters can be computed from <5 minutes signals (ultra-short-term analysis), ~5 minutes signals (short-term), and 24-hours signals (long-term). Accordingly, a window length is determined in our pipeline, from which HR and HRV parameters are extracted. In the proposed pipeline, we compute ultra-short-term and short-term HRV parameters, including AVNN, SDNN, RMSSD, pNN50, LF, HF, LFHF, SD1, SD2, and SD1SD2 [23].

IV. Results and Evaluation

We evaluate the proposed PPG pipeline using the dataset described in Section II. We employ data from 31 individuals, including 28736 minutes of PPG signals to train and validate ML models for signal quality assessment, noise reconstruction, and peak detection. The remaining data from 15 participants, encompassing 12288 minutes of PPG signals, are utilized to evaluate the accuracy of the pipeline (testing phase).

In our evaluation, we perform short-term and ultra-short-term HRV analyses. For the short-term analysis, we extract HR, AVNN, SDNN, RMSSD, pNN50, LF, HF, LFHF, SD1, SD2, and SD1SD2 measures from 5-minute signal windows. In contrast, the ultra-short-term analysis involves the extraction of HR, RMSSD, pNN50, and HF from 60-second signals [24]. To more effectively evaluate the performance of the proposed pipeline, we divide our test into two phases: one using the data collected during sleep and the other using the data collected during awake time. During sleep, users are typically still, and PPG signals are often cleaner and easier to analyze. However, during awake time, signals are more likely to be distorted due to hand movements and other sources of noise.

We use ECG signals as the gold reference in our analysis and apply our proposed pipeline to extract HR and HRV parameters from PPG signals. These values are similarly derived from the reference ECG signals using the Elgendi et al. method [25]. Then, we investigate the linear relationship between HR and HRV parameters derived from the PPG and ECG signals by employing linear regression analysis. In addition, we compute the Pearson correlation coefficient, mean absolute error (MAE), and ±95% confidence intervals of the differences between the pairwise HR and HRV parameters extracted from PPG and reference ECG.

A. Regression analysis of the proposed PPG pipeline

We perform a regression analysis to compare the HR and HRV parameters of PPG (obtained by the proposed pipeline) with HR and HRV parameters of reference ECG. Fig. 4 illustrates the regression analysis results of the 5-minute HR and HRV parameters. The blue and red lines in the plots represent the sleep and awake time data regression lines, while the black line (y=x) represents the optimal outcome when the parameters are identical. As indicated, the fitted lines for sleep data (in blue) closely align with the ideal line for all HR and

TABLE I: Statistical results of the comparison of 5-minute HR and HRV parameters between PPG and reference ECG

	Status	Method	HR	SDNN	RMSSD	pNN50	AVNN	LF	HF	LFHF	SD1	SD2	SD1SD2
Correlation	Sleep	Proposed	**0.99**	**0.99**	**0.99**	**0.98**	**0.99**	**0.85**	**0.86**	**0.92**	**0.99**	**0.99**	**0.97**
		Neurokit	0.98	0.98	0.96	0.96	**0.99**	0.78	0.78	0.89	0.96	0.97	0.92
		HeartPy	**0.99**	0.97	0.97	0.95	**0.99**	0.74	0.67	0.79	0.97	0.95	0.87
	Awake	Proposed	**0.99**	**0.99**	**0.94**	**0.96**	**0.99**	**0.80**	**0.83**	**0.74**	**0.94**	**0.98**	**0.79**
		Neurokit	0.85	0.94	0.76	0.79	0.94	0.78	0.47	0.31	0.76	0.94	0.44
		HeartPy	**0.99**	0.92	0.80	0.86	**0.99**	0.77	0.69	0.47	0.81	0.90	0.36
Mean-absolute -error	Sleep	Proposed	0.54	**2.59**	**5.53**	7.49	9.73	**3.57**	**4.88**	**0.21**	**3.92**	**2.64**	**0.05**
		Neurokit	0.69	3.53	7.68	**7.32**	**7.53**	4.36	6.49	0.28	5.44	4.20	0.07
		HeartPy	**0.43**	4.99	13.94	10.61	7.69	5.17	9.72	0.44	9.88	5.75	0.14
	Awake	Proposed	**0.77**	**2.93**	**9.21**	**4.71**	12.50	**3.65**	**3.07**	**0.54**	**6.53**	**2.42**	**0.10**
		Neurokit	2.37	6.86	27.93	10.67	23.37	8.32	16.83	1.32	19.83	6.77	0.28
		HeartPy	0.80	7.28	26.64	7.17	**8.83**	7.78	14.90	1.29	18.87	5.94	0.28
Confidence interval	Sleep	Proposed	[-1.6,2.7]	**[-7.5,5.4]**	**[-15,9.7]**	**[-3,17]**	[-58,39]	**[-10,11]**	**[-13,15]**	**[-0.65,0.78]**	**[-11,6.9]**	**[-8,6.5]**	**[-0.13,0.09]**
		Neurokit	[-3.8,5.1]	[-11,10]	[-25,15]	[-6.2,18]	**[-30,16]**	[-15,14]	[-22,18]	[-0.74,1.1]	[-18,11]	[-13,14]	[-0.22,0.12]
		HeartPy	**[-1.4,1.4]**	[-14,10]	[-32,7.5]	[-6.6,26]	[-26,26]	[-17,14]	[-32,18]	[-1,1.8]	[-23,5.3]	[-15,19]	[-0.34,0.09]
	Awake	Proposed	**[-1.8,2.7]**	**[-7,2.1]**	**[-22,5.2]**	**[-9.4,13]**	[-78,53]	**[-8.7,6.2]**	**[-9.3,5.1]**	**[-1.4,2.2]**	**[-16,3.7]**	**[-6.1,4]**	**[-0.26,0.07]**
		Neurokit	[-12,17]	[-20,9.9]	[-61,5.4]	[-30,13]	[-140,94]	[-26,11]	[-51,18]	[-1.7,4.2]	[-43,3.8]	[-16,21]	[-0.59,0.05]
		HeartPy	[-2.1,3.6]	[-18,7.1]	[-45,-8.5]	[-19,13]	**[-28,19]**	[-26,13]	[-43,13]	[-1.6,4.1]	[-32,-6]	[-18,20]	[-0.55,-0.01]

P-values < 0.0001

HRV parameters, except for pNN50, which shows a relatively higher degree of divergence. In awake time data, the regression lines (in red) closely align with the ideal line for HR, SDNN, RMSSD, AVNN, LF, HF, SD1, SD2, and SD1SD2. However, for pNN50 and LFHF, the fitted lines show a higher deviation.

Moreover, Fig. 5 shows the regression analysis for 60-second HR and HRV. Sleep data shows close alignment between the regression lines (in blue) of HR, RMSSD, and pNN50 with the ideal line. In contrast, HF indicates relatively higher deviation. Additionally, in awake time data, HR and pNN50 fitted lines (in red) closely align with the ideal line, while RMSSD and HF shows a higher degree of divergence.

In summary, our pipeline shows strong concordance with the reference ECG for most HR and HRV parameters in both sleep and awake time data.

B. Comparison with Existing Benchmark Methods

We compare our PPG pipeline with two existing benchmark methods: Neurokit [11] and HeartPY [13]. As an extension, we integrate a rule-based filter into these methods, similar to our pipeline's false peak detection filter, to discard PPG segments with more than 30% false peaks.

Table I shows the statistical results obtained from the proposed pipeline, Neurokit, and HeartPY methods in 5-minute segments. The results indicate that our proposed pipeline outperforms the other methods on sleep data in SDNN, RMSSD, pNN50, LF, HF, LFHF, SD1, SD2, and SD1SD2 with higher correlation, lower MAE, and narrower confidence intervals of the differences between the PPG and reference ECG. However, the proposed method and HeartPY obtained the best performance for HR, and all approaches performed similarly for AVNN. It is noteworthy that the proposed pipeline, Neurokit, and HeartPY approaches extracted 388, 388, and 351 sleep time 5-minute HRV samples from PPG signals, respectively.

During the subjects' wakeful states, our proposed pipeline demonstrates superior performance compared to the other approaches in terms of SDNN, RMSSD, pNN50, LF, HF, LFHF, SD1, SD2, and SD1SD2. However, no difference is observed between our method and HeartPy regarding HR and AVNN. It should be noted that the proposed pipeline, Neurokit, and HeartPY methods derived 97, 109, and 76 awake-time 5-minute HRV samples, respectively. Additionally, all methods have highly significant P-values below the threshold of 0.0001.

TABLE II: Statistical results of the comparison of 60-second HR and HRV parameters between PPG and reference ECG

	Status	Method	HR	RMSSD	pNN50	HF
Correlation	Sleep	Proposed	**0.96**	**0.86**	**0.84**	**0.48**
		Neurokit	0.94	0.83	0.83	0.41
		HeartPy	0.95	0.84	0.81	0.39
	Awake	Proposed	**0.90**	**0.72**	**0.80**	**0.60**
		Neurokit	0.78	0.52	0.55	0.25
		HeartPy	0.81	0.61	0.63	0.29
Mean-absolute -error	Sleep	Proposed	**1.99**	**13.84**	**10.88**	**21.01**
		Neurokit	2.13	15.63	11.16	23.62
		HeartPy	2.07	18.54	13.27	26.33
	Awake	Proposed	**2.96**	**13.89**	**7.53**	**19.52**
		Neurokit	3.63	32.56	15.15	41.15
		HeartPy	3.68	28.39	10.70	39.46
Confidence interval	Sleep	Proposed	**[-5.8,5.6]**	**[-41,35]**	**[-18,31]**	**[-57,56]**
		Neurokit	[-6.2,8.1]	[-46,38]	[-20,32]	[-66,59]
		HeartPy	[-5.5,6.6]	[-51,28]	[-17,36]	[-76,55]
	Awake	Proposed	**[-11,12]**	**[-41,17]**	**[-20,19]**	**[-62,44]**
		Neurokit	[-15,21]	[-79,16]	[-42,20]	[-120,49]
		HeartPy	[-13,17]	[-60,4.3]	[-29,21]	[-110,46]

P-values < 0.0001

Table II presents the statistical outcomes obtained from 60-second signal windows. As indicated, our method showed better performance compared to the other methods during sleep. The correlation coefficient for HF is relatively low, and its MAE and confidence interval values are found to be high in all methods. The proposed pipeline, Neurokit, and HeartPY methods extracted 2519, 2794, and 2433 sleep time 60-second HRV samples from PPG data, respectively.

In addition, our approach outperforms other methods on awake time 60-second data. Notably, Neurokit has demonstrated the worst performance with awake time data. The number of awake time 60-second HRV samples derived by the proposed pipeline, Neurokit, and HeartPY methods are 446, 980, and 820, respectively.

To summarize, the proposed PPG pipeline outperforms the other methods overall. In short-term analysis, our approach demonstrates better performance in both sleep and awake time in most HR and HRV parameters. In ultra-short-term analysis, while the three methods show similar results in sleep time, our method exhibits superior performance in awake time. These findings demonstrate our pipeline's effectiveness in reliably extracting HR and HRV parameters from PPG signals in both sleep and awake conditions.

V. Conclusion

Inaccurate HR and HRV measurements in PPG-based health monitoring systems can pose risks to patient safety, decision-making, and treatment effectiveness. In this paper, we proposed a robust end-to-end pipeline approach for reliable HR and HRV extraction from PPG signals. The proposed pipeline comprised three machine learning-based methods, starting with a signal quality assessment that discriminated between clean and noisy parts within signals. Then, a signal reconstruction technique reconstructed noisy parts by leveraging preceding patterns in the signal. Finally, a CNN-based approach identified systolic peaks in the signal and enabled the computation of HR and HRV parameters. The proposed PPG pipeline was evaluated on a dataset including PPG and ECG signals from 46 individuals recorded by smartwatch during daily activities. We also compared our approach with two established benchmark methods: Neurokit and HeartPY. The proposed PPG pipeline outperformed the other methods in terms of correlation and error between HR and HRV measures extracted from PPG and reference ECG.

References

[1] D. Castaneda et al. "A review on wearable photoplethysmography sensors and their potential future applications in health care". In: *International journal of biosensors & bioelectronics* 4.4 (2018), p. 195.

[2] J. Allen. "Photoplethysmography and its application in clinical physiological measurement". In: *Physiological measurement* 28.3 (2007), R1.

[3] D. Biswas et al. "Heart rate estimation from wrist-worn photoplethysmography: A review". In: *IEEE Sensors Journal* 19.16 (2019), pp. 6560–6570.

[4] Z. Zhang et al. "TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise". In: *IEEE. Trans. Biomed. Eng* 62.2 (2014), pp. 522–531.

[5] Z. Zhang. "Photoplethysmography-based heart rate monitoring in physical activities via joint sparse spectrum reconstruction". In: *IEEE transactions on biomedical engineering* 62.8 (2015), pp. 1902–1910.

[6] Y. Ye et al. "Combining nonlinear adaptive filtering and signal decomposition for motion artifact removal in wearable photoplethysmography". In: *IEEE Sensors Journal* 16.19 (2016), pp. 7133–7141.

[7] E. Grisan et al. "A supervised learning approach for the robust detection of heart beat in plethysmographic data". In: *2015 37th annual international conference of the IEEE EMBC*. IEEE. 2015, pp. 5825–5828.

[8] M. Essalat et al. "Supervised heart rate tracking using wrist-type photoplethysmographic (PPG) signals during physical exercise without simultaneous acceleration signals". In: *2016 IEEE GlobalSIP*. 2016, pp. 1166–1170.

[9] D. Biswas et al. "CorNET: Deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment". In: *IEEE Trans Biomed Circuits Syst* 13.2 (2019), pp. 282–291.

[10] K. Kazemi et al. "Robust PPG Peak Detection Using Dilated Convolutional Neural Networks". In: *Sensors* 22.16 (2022), p. 6054.

[11] Dominique M. et al. "NeuroKit2: A Python toolbox for neurophysiological signal processing". In: *Behavior Research Methods* 53.4 (2021), pp. 1689–1696.

[12] C. Carreiras et al. "Biosppy: Biosignal processing in python". In: *Accessed on* 3.28 (2015), p. 2018.

[13] P. Van Gent et al. "Heart rate analysis for human factors: Development and validation of an open source toolkit for noisy naturalistic heart rate data". In: *Proceedings of the 6th HUMANIST Conference*. 2018, pp. 173–178.

[14] F. Sarhaddi et al. "Trends in heart rate and heart rate variability during pregnancy and the 3-month postpartum period: Continuous monitoring in a free-living context". In: *JMIR mHealth and uHealth* (2022), e33458.

[15] I. Azimi et al. "Personalized maternal sleep quality assessment: An objective iot-based longitudinal study". In: *IEEE Access* 7 (2019), pp. 93433–93447.

[16] S. Jafarlou et al. "Objective prediction of next-day's affect using multimodal physiological and behavioral data: Algorithm development and validation study". In: *JMIR Formative Research* 7.1 (2023), e39425.

[17] F. Sarhaddi et al. "Long-term IoT-based maternal monitoring: system design and evaluation". In: *Sensors* 21.7 (2021), p. 2281.

[18] M. Mehrabadi et al. "Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study". In: *JMU* 8.11 (2020), e20465.

[19] *Gear Sport 42mm smartwatch*. https://www.samsung.com/us/mobile/wearables/smartwatches/gear-sport-blue-sm-r600nzbaxar/. Accessed April 2023.

[20] *Shimmer3 ECG Unit*. https://shimmersensing.com/product/shimmer3-ecg-unit-2/. Accessed April 2023.

[21] M. Feli et al. "An energy-efficient semi-supervised approach for on-device photoplethysmogram signal quality assessment". In: *Smart Health* (2023).

[22] Y. Wang et al. "Ppg signal reconstruction using deep convolutional generative adversarial network". In: *2022 44th Annual International Conference of the IEEE EMBC*. IEEE. 2022, pp. 3387–3391.

[23] F. Shaffer et al. "An overview of heart rate variability metrics and norms". In: *Frontiers in public health* (2017), p. 258.

[24] H. J. Baek et al. "Reliability of ultra-short-term analysis as a surrogate of standard 5-min analysis of heart rate variability". In: *Telemedicine and e-Health* 21.5 (2015), pp. 404–414.

[25] M. Elgendi et al. "Frequency bands effects on QRS detection". In: *Pan* 5.15 (2010), pp. 1–5.