Feature Stability of PPG Morphological Descriptors Across Diverse Clinical Acquisition Settings

Matteo Ricci[1,2], Lorenzo Fácila[3], Raúl Alcaraz[4], José J. Rieta[1], Frida Sandberg[2]
[1]Electronic Engineering Department, Universitat Politècnica de València, València, Spain
[2] Department of Biomedical Engineering, Lund University, Lund, Sweden
[3] Cardiology Department, General University Hospital Consortium of Valencia, Spain
[4] Research Group in Electronic, Biomed. and Telecomm. Eng., Univ. of Castilla-La Mancha, Spain

Abstract

Photoplethysmography (PPG) is a promising non-invasive technique for blood pressure (BP) estimation, although machine learning models often show limited generalizability across datasets due to feature variability. This study investigates the stability of PPG morphological features across recordings and cohorts to identify descriptors with consistent behavior. Analyses were performed on representative subsets from four public datasets: PPG-BP (219 subjects), VitalDB (285 subjects), MIMIC-III (225 subjects), and MIMIC-IV (63 subjects). First, stationary 30 s segments with minimal heart-rate variation were extracted from VitalDB, MIMIC-III, and MIMIC-IV. Then, 2.1 s segments were selected to enable comparison with the PPG-BP dataset. A total of 41 morphological features were extracted, and their robustness was assessed through a variability index, $\gamma_f^{(j)}$. Features below duration-specific thresholds (0.30 for 30 s and 0.13 for 2.1 s) were classified as robust. The number of robust features was consistent across datasets for the 2.1 s segments (19–20 features), whereas it ranged from 15 to 25 for the 30 s segments. Timing parameters (e.g., Δn_{up} and ΔT_{12}) and selected amplitude ratios (e.g., PDA A_{13}) showed the highest stability, whereas second-derivative ratios and dicrotic-notch metrics were highly variable. Prioritizing robust descriptors may improve the generalizability of machine learning models for BP estimation.

1. Introduction

Photoplethysmography (PPG) is a low-cost, widely accessible optical technique that has emerged as a promising non-invasive method for continuous blood pressure (BP) estimation [1]. The morphology of the peripheral pulse wave carries key information on cardiovascular function, enabling its use in diagnostic algorithms [2]. Accordingly, many studies extract features from the PPG to train models for accurate systolic and diastolic BP estimation [3, 4].

Despite their potential, machine learning models for BP estimation often exhibit limited generalization across populations [5, 6]. Clinical acquisition settings range from controlled ambulatory environments to intensive care units (ICUs), where patients may exhibit severe hemodynamic instability and receive pharmacological treatments [7]. This heterogeneity introduces multiple confounding factors, including increased signal noise and motion artifacts, which alter PPG morphology and challenge the robustness of commonly used features [8].

To address this issue, this study evaluates the within-segment stability of a comprehensive set of PPG morphological features acquired under stationary conditions across four public datasets. By quantifying feature stability under diverse clinical conditions, we aim to identify a subset of descriptors that remain robust across populations [9, 10]. Identifying these robust features is key to reducing dataset-specific bias and improving the generalizability and clinical reliability of non-invasive BP estimation models [6, 11].

2. Materials and Methods

Analyses were conducted on representative subsets from four clinical cohorts to capture different physiological states and acquisition conditions. The ambulatory population was represented by the PPG-BP dataset [12] (219 subjects), intraoperative recordings by VitalDB [13] (285 subjects), and critically ill patients by MIMIC-III-Ext-PPG [14] (225 subjects) and MIMIC-IV [15] (63 subjects).

All raw PPG signals were preprocessed using a 0.5–12 Hz IIR bandpass filter to remove baseline drift and high-frequency noise.

To account for the different native signal lengths, the analysis was divided into two phases. In the first phase, three 30-s segments per subject were extracted from VitalDB, MIMIC-III, and MIMIC-IV. Segments were selected under stationary conditions using three criteria: (i) heart rate between 40 and 180 bpm; (ii) cardiac spectral dominance, defined as a ratio > 0.8 between the power

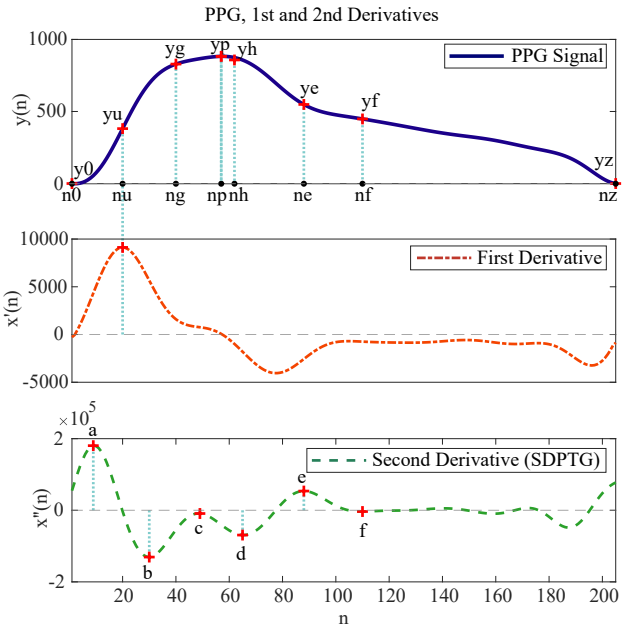

Figure 1: Fiducial-point detection in the PPG signal $y(n)$, first-derivative $x'(n)$, and second-derivative $x''(n)$.

in the 0.5–4 Hz band and the total signal power; and (iii) rhythmic stability with an interbeat-interval coefficient of variation < 0.10.

In the second phase, standardized 2.1-s subsegments were extracted from the validated 30-s segments of VitalDB, MIMIC-III, and MIMIC-IV, while the PPG-BP dataset, whose recordings are natively short, was used directly at this segment length. Segments were retained only if they contained at least three complete pulse waves and had a heart rate within physiological limits.

Segments for which more than 20% of the morphological features were non-computable were excluded. A total of 41 morphological descriptors were then computed from each segment and its pulse waves, including timing parameters, amplitude metrics, second-derivative indices (SDPTG a–e), and Pulse Decomposition Analysis (PDA) descriptors (Figures 1 and 2) [7, 16].

To quantify the relative intra-segment variability of feature f in dataset j, the variability index $\gamma_f^{(j)}$ is defined as

$$\gamma_f^{(j)} = \frac{\frac{1}{N_j}\sum_{i=1}^{N_j} \sigma_f^{(i,j)}}{\sigma_f^{(j)}},$$

where N_j is the number of extracted segments, $\sigma_f^{(i,j)}$ is the standard deviation of feature f within the ith segment, and $\sigma_f^{(j)}$ is the overall standard deviation of the same feature across the dataset.

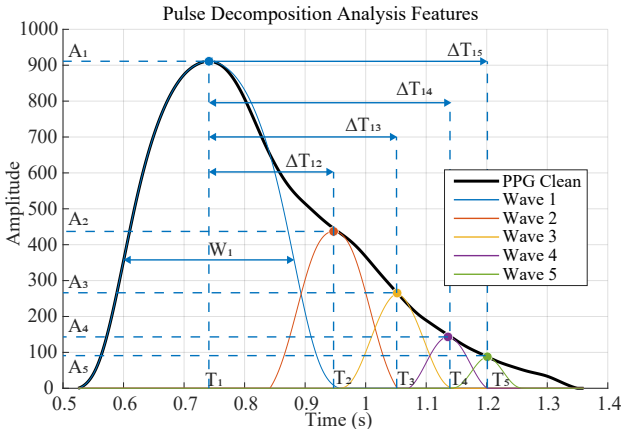

Figure 2: Feature extraction using Pulse Decomposition Analysis, modeling the pulse as the sum of five component waves with amplitudes (A_i) and temporal intervals (ΔT_{ij}).

The final variability index γ_f was computed as the median value across datasets.

3. Results

The distribution of γ_f in the 30-s and 2.1-s segments is visualized in Figure 3. Furthermore, the late-diastolic PDA features ΔT_{14}, A_{14}, ΔT_{15}, and A_{15} were systematically excluded due to fitting instability, as they generated missing values in more than 60% of the accepted segments. Their removal left 37 features for evaluation; the remaining features were computable in more than 92% of the accepted segments.

An in-depth evaluation of individual parameter performance highlighted distinct robustness profiles across the two segment lengths. In the 30-s segments, the lowest cross-dataset variability (γ_f) was obtained for the fiducial-point time interval Δn_{up} (0.118, systolic upstroke), followed by PDA ΔT_{12} (0.135, reflection delay). Conversely, the highest γ_f was obtained for the second-derivative ratio c/a (0.466, late systolic inflection), whereas other SDPTG parameters showed similarly high variability. In the 2.1-s segments, the lowest γ_f was obtained for the systolic peak amplitude y_p (0.049, peak-to-peak height), followed by heart rate (HR; 0.060). In contrast, the highest γ_f was observed for the SDPTG ratio d/a (0.215, late systolic slope inflection), followed closely by other derivative ratios such as b/a and c/a.

PPG features were considered robust when their $\gamma_f^{(j)}$ remained below a predefined threshold. Specifically, thresholds of 0.30 and 0.13 were adopted for the 30-s and 2.1-s segments, respectively; each was calibrated at the 60th percentile of the overall γ_f distribution to retain the 60% least variable features, thereby facilitating comparison between

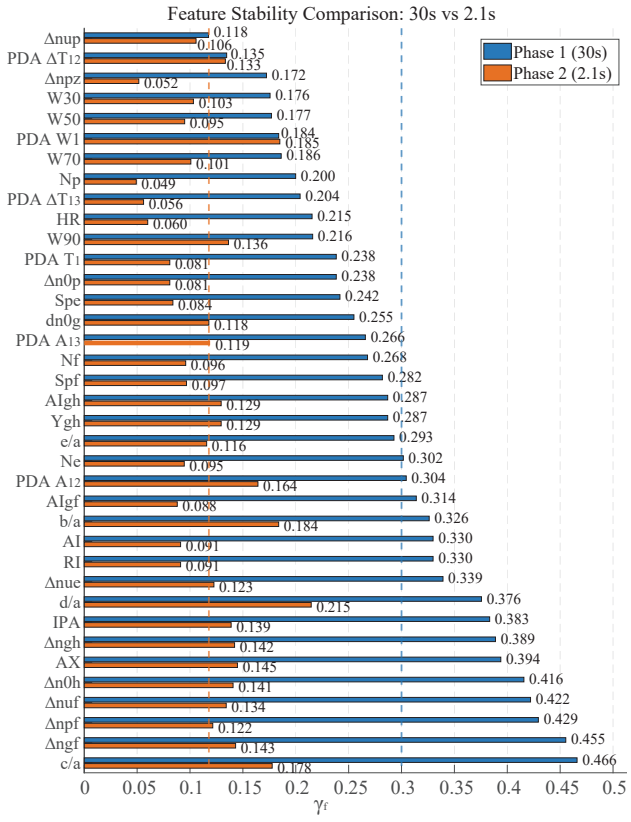

Figure 3: The plot compares feature robustness between Phase 1 (30 s) and Phase 2 (2.1 s) using the median variability index (γ_f) across datasets. Features are ranked according to γ_f, while the vertical dashed lines indicate the variability thresholds (0.30 and 0.13, respectively).

the two cases.

Feature robustness varied across datasets and segment lengths. For the 2.1-s segments, the ambulatory PPG-BP dataset presented 19 robust features. Similar results were obtained in the clinical datasets, with 20 robust features identified in VitalDB, 19 in MIMIC-III, and 20 in MIMIC-IV. For the 30-s segments, 23 robust features were identified in VitalDB, 15 in MIMIC-III, and 25 in MIMIC-IV. As shown in Table 1, robustness varied across feature categories. PDA features demonstrated substantial variation in robustness rates, ranging from 16.7% in VitalDB (2.1 s) to 66.7% in several dataset configurations.

SDPTG feature robustness varied considerably across datasets, with rates between 20.0% and 80.0% in the clinical datasets and 100.0% in PPG-BP. As previously noted, specific markers within this family, such as the d/a and c/a ratios, showed high intra-segment variability. Features derived from the PPG waveform and its first derivative represented the largest group and maintained a consistent robustness rate in 2.1-s segments (46.2% to 61.5%), with similar rates across 30-s segments (42.3% to 65.4%).

Table 1: Robust Morphological Features Across Datasets and Segment Lengths

Dataset	PPG & 1st Deriv	SDPTG (2nd Deriv)	PDA Features
Total evaluated	26	5	6
30-s segments (Target: 60th Percentile)			
VitalDB	17 (65.4%)	3 (60.0%)	3 (50.0%)
MIMIC-III	11 (42.3%)	1 (20.0%)	3 (50.0%)
MIMIC-IV	17 (65.4%)	4 (80.0%)	4 (66.7%)
2.1-s segments (Target: 60th Percentile)			
PPG-BP	12 (46.2%)	5 (100.0%)	2 (33.3%)
VitalDB	16 (61.5%)	3 (60.0%)	1 (16.7%)
MIMIC-III	13 (50.0%)	2 (40.0%)	4 (66.7%)
MIMIC-IV	14 (53.8%)	3 (60.0%)	3 (50.0%)

4. Discussion

Features derived from the PPG waveform and its first derivative form the most consistently robust group; specifically, features such as the systolic upstroke time (Δn_{up}) and N_p show high stability across cohorts and have known associations with BP [17]. Conversely, temporal distances from the dicrotic notch (e.g., Δn_{gf}) and second-derivative ratios (c/a, b/a, and d/a) exhibit high intra-segment variability.

Among PDA features, robustness depends on segment length: the systolic-to-reflected peak interval (ΔT_{12}) is highly stable in 30-s windows [18] but varies in 2.1-s segments, where ΔT_{13} is more robust.

However, two limitations remain. First, differing segment durations inherently affect variance comparisons despite threshold calibration. Second, feature stability does not guarantee predictive utility and therefore requires further validation within a complete cross-dataset BP estimation framework.

5. Conclusion

This cross-dataset evaluation demonstrates that PPG feature robustness depends on the dataset and segment length, although the number of robust features was consistent across datasets for the 2.1-s segments. Prioritizing morphological descriptors with demonstrated cross-dataset stability, particularly timing features from the PPG waveform and its first derivative, may improve the generalizability of machine learning models for cuffless BP estimation.

Acknowledgments

This research received partial financial support from Spanish Government grants PID2021-123804OB-I00, PID2021-128525OB-I00, and PID2025-73302OB-I00.

These grants were jointly funded by the European Regional Development Fund (EU; funder identifier 10.13039/501100011033). Additional support was provided by grant SBPLY/24/180225/000184 from the Junta de Comunidades de Castilla-La Mancha and grant CIAICO/2023/200 from the Generalitat Valenciana.

References

[1] Tang H, Ma G, Qiu L, Zheng L, Bao R, Liu J, Wang L. Blood Pressure Estimation Based on PPG and ECG Signals Using Knowledge Distillation. Cardiovascular Engineering and Technology February 2024;15(1):39–51. ISSN 1869-4098.

[2] Elgendi M. On the Analysis of Fingertip Photoplethysmogram Signals. Current Cardiology Reviews February 2012; 8(1):14–25. ISSN 1573-403X.

[3] Xing X, Sun M. Optical blood pressure estimation with photoplethysmography and FFT-based neural networks. Biomedical Optics Express July 2016;7(8):3007–3020. ISSN 2156-7085.

[4] Attivissimo F, De Palma L, Di Nisio A, Scarpetta M, Lanzolla AML. Photoplethysmography Signal Wavelet Enhancement and Novel Features Selection for Non-Invasive Cuff-Less Blood Pressure Monitoring. Sensors January 2023;23(4):2321. ISSN 1424-8220.

[5] Moulaeifard M, Charlton PH, Strodthoff N. Generalizable deep learning for photoplethysmography-based blood pressure estimation – A Benchmarking Study, February 2025. ArXiv:2502.19167 [cs].

[6] Mehta S, Kwatra N, Jain M, McDuff D. Examining the challenges of blood pressure estimation via photoplethysmogram. Scientific Reports August 2024;14(1):18318. ISSN 2045-2322.

[7] Weber-Boisvert G, Gosselin B, Sandberg F. Intensive care photoplethysmogram datasets and machine-learning for blood pressure estimation: Generalization not guarantied. Frontiers in Physiology March 2023;14:1126957. ISSN 1664-042X.

[8] Caizzone A, Boukhayma A, Enz C. Comprehensive noise analysis in PPG read-out chains. In 2017 International Conference on Noise and Fluctuations (ICNF). June 2017; 1–4.

[9] Xu Y, He Z, Wang H. Multi-Source Causal Invariance for Cuffless Blood Pressure Estimation Based on Photoplethysmography Signal Features. Sensors January 2025; 25(11):3254. ISSN 1424-8220.

[10] Mohagheghian F, Han D, Peitzsch A, Nishita N, Ding E, Dickson EL, DiMezza D, Otabil EM, Noorishirazi K, Scott J, Lessard D, Wang Z, Whitcomb C, Tran KV, Fitzgibbons TP, McManus DD, Chon KH. Optimized Signal Quality Assessment for Photoplethysmogram Signals Using Feature Selection. IEEE Transactions on Biomedical Engineering September 2022;69(9):2982–2993. ISSN 1558-2531.

[11] Huthart S, Elgendi M, Zheng D, Stansby G, Allen J. Advancing PPG Signal Quality and Know-How Through Knowledge Translation-From Experts to Student and Researcher. Frontiers in Digital Health 2020;2:619692. ISSN 2673-253X.

[12] Liang Y, Chen Z, Liu G, Elgendi M. A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. Scientific Data February 2018;5(1):180020. ISSN 2052-4463.

[13] Lee HC, Jung CW. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients.

[14] Moulaeifard M, Charlton PH, Strodthoff N. MIMIC-III-Ext-PPG: A PPG Benchmark Dataset for Cardiorespiratory Analysis.

[15] Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, Pollard TJ, Hao S, Moody B, Gow B, Lehman LwH, Celi LA, Mark RG. MIMIC-IV, a freely accessible electronic health record dataset. Scientific Data January 2023;10(1):1. ISSN 2052-4463.

[16] Marc-Derrien Y, Gren L, Dierschke K, Albin M, Gudmundsson A, Wierzbicka A, Sandberg F. Acute Cardiovascular Effects of Hydrotreated Vegetable Oil Exhaust. Frontiers in Physiology 2022;13:828311. ISSN 1664-042X.

[17] Chowdhury MH, Shuzan MNI, Chowdhury ME, Mahbub ZB, Uddin MM, Khandakar A, Reaz MBI. Estimating Blood Pressure from the Photoplethysmogram Signal and Demographic Features Using Machine Learning Techniques. Sensors Basel Switzerland June 2020;20(11):3127. ISSN 1424-8220.

[18] Huang CH, Guo JW, Yang YC, Tsai PY, Wu AYA, Lin HJ, Wang TD. Weighted Pulse Decomposition Analysis of Fingertip Photoplethysmogram Signals for Blood Pressure Assessment. In 2020 IEEE International Symposium on Circuits and Systems (ISCAS). ISSN 2158-1525, October 2020; 1–5.

Address for correspondence:

José J. Rieta

BioMIT.org, Electronic Engineering Department, Building 7F, Universitat Politècnica de València, 46022 Valencia, Spain.

e-mail: jjrieta@upv.es