Generalizable PPG Signal Quality Assessment Using Correlation-Derived Morphological Features

Pietro Chierico[1], Vicente Bertomeu-González[2], Raúl Alcaraz[3], Mohamed Elgendi[4], José J Rieta[1]

[1] BioMIT.org, Electronic Engineering Department, Universitat Politecnica de Valencia, Spain
[2] Department of Clinical Medicine, Miguel Hernández University, Spain
[3] Research Group in Electronic, Biomed. and Telecomm. Eng., Univ. of Castilla-La Mancha, Spain
[4] Dep. of Biomedical Eng. and Biotechnology, Khalifa University, Abu Dhabi, United Arab Emirates

Abstract

Continuous monitoring of photoplethysmographic (PPG) signals enables estimation of physiological parameters such as heart rate and pulse rate variability, but signal quality is often degraded by motion artifacts and environmental noise, limiting reliability. This study proposes an interpretable binary signal quality assessment (SQA) method based on morphological self-consistency using correlation analysis and machine learning. PPG signals were segmented into fixed-length windows, and beat-centered segments were compared using sliding Pearson correlation to derive three low-dimensional descriptors capturing waveform consistency and dispersion. These features were used to train a neural network classifier on the MIMIC-III-Ext-PPG dataset and evaluated on external wearable datasets, achieving accuracies of 87.1%, 88.4%, 89.9%, and 82.9% on MIMIC-III, PPG-DaLiA, WESAD, and TROIKA, respectively. The results support the robustness and practical relevance of correlation-derived morphological features for wearable PPG quality control.

1. Introduction

Continuous monitoring of photoplethysmographic (PPG) signals enables estimation of physiological parameters such as heart rate (HR), pulse rate variability (PRV), and blood pressure (BP) [1]. However, signal quality is often degraded by motion artifacts, poor sensor contact, and environmental noise, which can significantly compromise the reliability of derived measurements [2, 3]. Despite extensive research, there is no standardized approach for automatic PPG signal quality assessment (SQA), and many existing methods rely on fiducial point detection or handcrafted signal quality indices (SQIs) [4–6], making them sensitive to noise and algorithm-dependent.

In this study, we propose a novel SQA framework based on morphological self-consistency analysis using correlation. The method evaluates beat-to-beat waveform similarity through sliding Pearson correlation of beat-centered segments, extracting features that capture intrinsic morphological stability. These features are mapped into a compact representation and used to train a machine learning classifier for automatic binary quality discrimination. Unlike conventional approaches, the proposed method does not rely on handcrafted SQIs or a large feature pool. Compared with our previous CinC work, which relied on a broader handcrafted feature set and feature selection, the present study focuses on a compact and interpretable representation based on morphological self-consistency, with explicit evaluation across heterogeneous datasets. The framework is therefore designed to improve robustness, interpretability, and generalization in wearable monitoring applications.

2. Materials and Methods

This study aimed to characterize photoplethysmographic (PPG) signal quality through a compact feature representation derived from pulse morphology consistency. Two complementary data sources were used for model development and evaluation. For internal development, a subset of 1,962 PPG segments was extracted from the MIMIC-III-Ext-PPG database [7], a publicly available resource derived from the MIMIC-III Waveform Database Matched Subset. This dataset contains high-quality ICU-acquired PPG signals, typically sampled at 125 Hz, along with rhythm annotations and, when available, synchronized physiological waveforms. To enhance generalizability and reduce subject bias, the analysis was restricted to 100 patients, selecting 3 minutes and 20 seconds of data per subject, thus ensuring a controlled and representative training setting. For external evaluation, the Segade dataset [8] was used, a labeled dataset for PPG artifact detection comprising 21,921 signal segments from 42 subjects. Signals are annotated at the sample level with binary labels indicating the presence or absence of artifacts (artifact vs non-artifact). The Segade dataset aggregates three publicly available wrist-worn datasets (PPG-DaLiA, WESAD, and TROIKA), providing a heterogeneous evaluation scenario with varying noise and motion conditions. Given the sample-wise annotation scheme, the data were reorganized by grouping and concatenating samples sharing the same label, enabling the construction of fixed-length segments with consistent quality annotations.

This dual-dataset setup enables a comprehensive evaluation of the proposed method on both clean, clinically acquired signals and noisy, motion-corrupted wearable data, allowing assessment of robustness and generalization across heterogeneous acquisition conditions.

The overall pipeline was implemented in MATLAB (R2023b, MathWorks, Natick, MA, USA) and applied independently to each signal segment. Each raw PPG recording was segmented into fixed-length windows of 10 seconds. This duration represents a trade-off between temporal resolution and the need to include a sufficient number of cardiac cycles for reliable morphology comparison [4].

The proposed method quantifies morphological self-consistency within each segment by evaluating beat-to-beat similarity. For each detected pulse, a peak-centered window is extracted using a local-extrema strategy based on the MATLAB *findpeaks* function. Each template is then compared with the rest of the segment using sliding Pearson correlation, generating a set of correlation curves.

These curves are assembled into a matrix and converted into a pairwise Pearson correlation matrix that captures internal waveform agreement. As illustrated in Figure 1, high-quality segments tend to produce coherent correlation patterns, whereas noisy segments generate more heterogeneous structures.

From this representation, low-dimensional descriptors are extracted to summarize the consistency and dispersion of pulse morphology. Specifically, the mean and standard deviation of the correlation statistics are projected into a two-dimensional point cloud, from which the centroid coordinates and the area of a confidence ellipse are computed. These three scalar features provide a compact description of segment quality, where compact ellipses and stable centroids indicate more regular morphology and higher internal consistency.

The extracted features were stored at the segment level together with the binary quality label, thus creating a supervised dataset for downstream classification. This results in a computationally efficient and transparent pipeline suitable for evaluation across heterogeneous datasets. Model development and evaluation were performed using a subject-wise nested cross-validation (CV) strategy. To prevent data leakage, all segments belonging to the same subject were assigned to a single fold. The

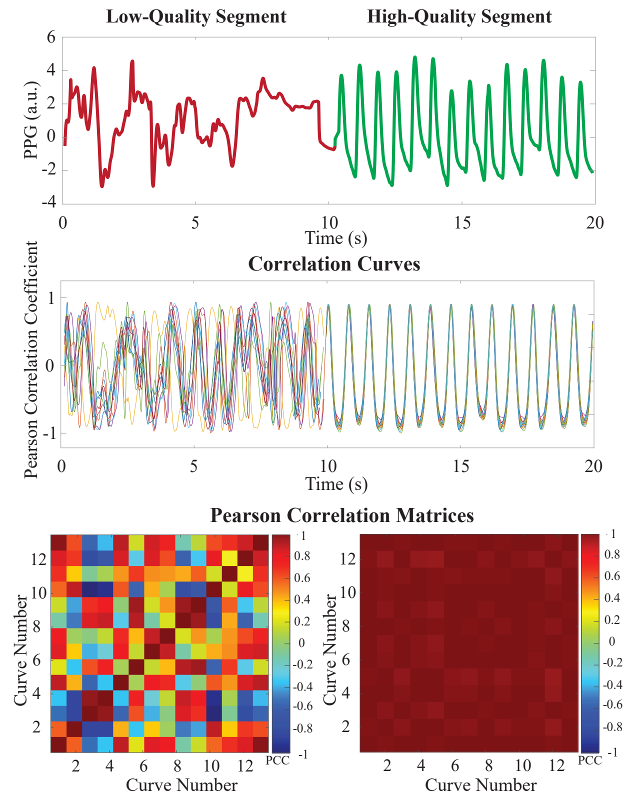

Figure 1. Illustration of the proposed correlation-based PPG signal quality assessment. Top: high-quality (right) and low-quality (left) segments. Middle: corresponding Pearson correlation curves, showing consistent patterns for high-quality signals and irregular behavior for low-quality signals. Bottom: pairwise correlation matrices, with a compact structure for high-quality segments and a heterogeneous pattern for low-quality ones. PCC in this case stands for Pearson Correlation Coefficient.

nested design consisted of an outer loop for unbiased performance estimation and an inner loop dedicated to hyperparameter tuning. Within the inner loop, model hyperparameters were optimized, and the selected configuration was subsequently evaluated on the corresponding outer test fold. The use of a MIMIC-derived subset ensured exposure to clinically diverse waveform patterns, while the external dataset enabled testing under wearable and motion-corrupted conditions. Within this setup, the proposed method was evaluated alongside established signal quality indices (SQIs) to enable a consistent comparison with existing approaches.

Table 1. Classification performance (mean $\pm$ standard deviation) obtained via subject-wise nested cross-validation. The reported configuration for MIMIC-III corresponds to the final selected model (LayerSizes = 5, $\lambda = 0.001$).

Dataset	Acc	Rec	F1
MIMIC-III	0.8707 ± 0.0189	0.9670 ± 0.0162	0.9265 ± 0.0108
PPG-DaLiA	0.8836 ± 0.0028	0.8737 ± 0.0975	0.8790 ± 0.0331
WESAD	0.8989 ± 0.0033	0.8987 ± 0.0323	0.8988 ± 0.0042
TROIKA	0.8294 ± 0.0210	0.8299 ± 0.0122	0.8292 ± 0.0085

3. Results

The classification performance of the proposed method across the considered datasets is reported in Table 1, with results presented as mean $\pm$ standard deviation over repeated runs. Among the evaluated machine learning models, a Narrow Neural Network architecture was selected, with its configuration determined through the inner loop of the subject-wise nested cross-validation procedure.

The model consists of a single fully connected hidden layer, where the number of neurons (LayerSizes) and the regularization parameter λ were treated as hyperparameters and optimized during training. Across the nested folds, the selected configurations varied, with LayerSizes ranging between 5 and 15 and λ spanning from 0 to 10^{-3}. Based on this procedure, the final model configuration was defined by a hidden layer size of 5 neurons and a regularization parameter $\lambda = 0.001$.

Using this configuration, the proposed method achieves robust and consistent performance across all datasets. In particular, the model attains an accuracy of 0.8707 ± 0.0189 on MIMIC-III, 0.8836 ± 0.0028 on PPG-DaLiA, 0.8989 ± 0.0033 on WESAD, and 0.8294 ± 0.0210 on TROIKA.

The observed differences in performance across datasets are consistent with the expected variability in acquisition conditions. MIMIC-III provides relatively clean clinical signals, whereas wearable datasets such as PPG-DaLiA, WESAD, and TROIKA include motion artifacts, sensor displacement, and environmental noise. Despite this, the method maintains stable performance, as indicated by the low standard deviations across folds.

Importantly, the performance decrease observed in more challenging datasets such as TROIKA reflects increased signal heterogeneity rather than model instability. The method remains robust under these conditions, suggesting that the proposed representation captures intrinsic morphological properties of the PPG signal that generalize across acquisition domains.

Although a neural network was selected for its balance between computational efficiency and nonlinear modeling capability, the low dimensionality and structured nature of the proposed feature space suggest that the classification task is not strongly model-dependent. This is supported by preliminary experiments with alternative classifiers, such as Support Vector Machines and k-Nearest Neighbors, which yielded comparable results, indicating that performance is primarily driven by the morphological self-consistency representation rather than the specific classifier.

4. Discussion

The proposed approach demonstrates that a compact representation of morphological self-consistency is sufficient to achieve competitive performance in PPG signal quality assessment. Unlike traditional approaches that rely on multiple handcrafted features or precise fiducial detection, the proposed method captures signal quality through internal waveform agreement, providing a more intrinsic characterization of signal reliability.

From a physiological perspective, the results can be interpreted in terms of waveform stability. High-quality PPG signals are characterized by consistent pulse morphology, reflecting stable cardiac cycles and adequate peripheral perfusion. In contrast, low-quality segments exhibit increased variability due to motion artifacts, poor sensor contact, or transient physiological changes [9]. The correlation-based representation effectively captures this distinction by quantifying the degree of agreement between pulse shapes within each segment.

The comparison with classical signal quality indices (SQIs), including perfusion index (PSQI), skewness (SSQI), and kurtosis (KSQI) [5], further highlights the advantages of the proposed method. While SQI-based approaches achieve comparable performance on controlled datasets such as MIMIC-III, their performance degrades significantly in wearable datasets, as shown in Table 2. This suggests that single-feature descriptors are insufficient to capture the complexity of real-world PPG signals.

In contrast, the proposed morphology-consistency representation integrates multiple aspects of signal behavior into a unified framework. The use of correlation curves allows capturing both local waveform similarity and global structural consistency, resulting in improved robustness under varying acquisition conditions. Additionally, the low dimensionality of the feature space (centroid coordinates and ellipse area) ensures computational efficiency and interpretability.

Another important aspect is the reduced dependence on the specific classification model. Preliminary experiments with alternative classifiers (e.g., SVM and k-NN) yielded similar results, indicating that performance is primarily driven by the feature representation rather than the learning algorithm. This property is particularly relevant for practical applications, where simpler models may be preferred for real-time implementation.

Table 2. Accuracy comparison between the proposed method and SQI-based approaches across datasets

Dataset	Proposed	PSQI	SSQI	KSQI
MIMIC-III	0.8707	0.8571	0.8627	0.8515
PPG-DaLiA	0.8836	0.4294	0.5205	0.6425
WESAD	0.8989	0.5085	0.5943	0.7622
TROIKA	0.8294	0.4734	0.5740	0.5414

Despite these advantages, some limitations should be noted. First, the lack of standardized and large-scale annotated datasets for PPG signal quality assessment remains a challenge. Differences in annotation criteria and acquisition setups can introduce variability in evaluation. Second, although the proposed method avoids explicit reliance on fiducial detection, it still depends on peak detection for segment extraction, which may be affected under extreme noise conditions.

Future work should focus on evaluating the method on larger and more diverse datasets, as well as exploring its integration into real-time wearable systems. Additionally, extending the framework to multi-level quality assessment or continuous quality scoring could further enhance its applicability.

Overall, the results support the use of morphological self-consistency as a robust and generalizable descriptor for PPG signal quality assessment, particularly in heterogeneous and real-world monitoring scenarios.

5. Conclusions

The proposed method quantifies morphological self-consistency through a compact set of correlation-derived morphological descriptors. It achieved robust performance across heterogeneous datasets while remaining computationally efficient. Its simplicity, interpretability, and resilience to noise make it a promising candidate for real-time quality control in wearable monitoring systems.

Acknowledgments

Research funded by grants PID2021-123804OB-I00, PID2021-00X128525-IV0 and TED2021-130935B-I00 from the Spanish Government and the European Regional Development Fund (10.13039/501100011033), SBPLY/21/180501/000186 from Junta de Castilla-La Mancha and AICO/2023/200 from Generalitat Valenciana.

References

[1] Allen J, Zheng D, Kyriacou PA, Elgendi M. Photoplethysmography (ppg): state-of-the-art methods and applications, 2021.

[2] Charlton PH, Marozas V, Mejía-Mejía E, Kyriacou PA, Mant J. Determinants of photoplethysmography signal quality at the wrist. PLOS Digital Health 2025;4(6):e0000585.

[3] Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiological measurement March 2007;28(3):R1–39. ISSN 0967-3334.

[4] Orphanidou C. Quality assessment for the photoplethysmogram (ppg). In Signal Quality Assessment in Physiological Monitoring: State of the Art and Practical Considerations. Springer, 2017; 41–63.

[5] Elgendi M. Optimal signal quality index for photoplethysmogram signals. Bioengineering 2016;3(4):21.

[6] Moscato S, Palmerini L, Palumbo P, Chiari L. Quality assessment and morphological analysis of photoplethysmography in daily life. Frontiers in Digital Health 2022;4:912353.

[7] Moulaeifard M, Charlton PH, Strodthoff N. MIMIC-III-Ext-PPG: A PPG Benchmark Dataset for Cardiorespiratory Analysis. PhysioNet March 2026;Version 1.1.0.

[8] Guo Z, Ding C, Hu X, Rudin C. A supervised machine learning semantic segmentation approach for detecting artifacts in plethysmography signals from wearables. Physiological Measurement 2021;42(12):125003.

[9] Dendorfer A, Charlton PH. Wrist photoplethysmography pulse waves: Morphological classes and physiological influences. Engineering Proceedings 2025;118(1):83.

Address for correspondence:

José J. Rieta

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

e-mail: jjrieta@upv.es