Self-Supervised and Topological Signal-Quality Assessment for Any PPG Device

Wei Shao[1], Ruoyu Zhang[2], Zequan Liang[1], Ehsan Kourkchi[2], Setareh Rafatirad[1], and Houman Homayoun[2]

[1]Department of Computer Science, University of California, Davis, Davis, CA, U.S.A.
[2]Department of Electrical and Computer Engineering, University of California, Davis, Davis, CA, U.S.A.
Email: {wayshao, ryuzhang, zqliang, ekay, srafatirad, hhomayoun}@ucdavis.edu

***Abstract*—Wearable photoplethysmography (PPG) is embedded in billions of devices, yet its optical waveform is easily corrupted by motion, perfusion loss, and ambient light—jeopardizing downstream cardiometric analytics. Existing signal-quality assessment (SQA) methods rely either on brittle heuristics or on data-hungry supervised models. We introduce the first fully unsupervised SQA pipeline for wrist PPG. Stage 1 trains a contrastive 1-D ResNet-18 on 276 h of raw, unlabeled data from heterogeneous sources (varying in device and sampling frequency), yielding optical-emitter– and motion-invariant embeddings (i.e., the learned representation is stable across differences in LED wavelength, drive intensity, and device optics, as well as wrist motion). Stage 2 converts each 512-D encoder embedding into a 4-D topological signature via persistent homology (PH) and clusters these signatures with HDBSCAN. To produce a binary signal-quality index (SQI), the acceptable PPG signals are represented by the densest cluster while the remaining clusters are assumed to mainly contain poor-quality PPG signals. Without re-tuning, the SQI attains Silhouette, Davies–Bouldin, and Calinski–Harabasz scores of 0.72, 0.34, and 6,173, respectively, on a stratified sample of 10,000 windows. In this study, we propose a hybrid self-supervised-learning–topological-data-analysis (SSL–TDA) framework that offers a drop-in, scalable, cross-device quality gate for PPG signals.**

***Index Terms*—photoplethysmography, signal quality, self-supervised learning, persistent homology, wearable sensing**

I. Introduction

Wearable photoplethysmography (PPG) underpins today's cardiometric ecosystem—delivering heart rate, SpO_2, respiration, and nascent cuff-less blood-pressure estimates in smartwatches, rings, and earbuds. Global shipments already exceed millions of units per year, generating petabyte-scale PPG streams. Yet the optical waveform is notoriously fragile: motion artifacts, ambient-light leakage, skin–sensor decoupling, and perfusion changes routinely degrade signal quality [1]–[3]. Without timely filtering, downstream algorithms can yield grossly erroneous vitals, undermining user trust and clinical adoption.

Commercial firmware embeds hand-tuned signal-quality assessment (SQA) heuristics—thresholds on amplitude, template correlation, or derivative energy—engineered per LED wavelength and mechanical stack; a firmware update or strap relocation can break these rules. Supervised CNNs detect artifacts reliably [4], but each hardware generation demands thousands of freshly labeled windows, rendering cross-device scaling impractical.

Wearables already store hundreds of hours of unlabeled wrist-PPG per user. Contrastive self-supervised learning (SSL) can harness this free data, but SSL alone does not output a human-interpretable signal-quality index (SQI). Conversely, topology-based descriptors capture waveform morphology in a few numbers, yet they have never been paired with modern deep encoders. Persistent homology (PH) has characterized cardiac periodicity and gait regularity [5]; to our knowledge, we are the first to use PH as a morphology prior for wrist-PPG quality.

We fuse SSL and topological data analysis (TDA) into the first fully unsupervised, device-agnostic SQA pipeline, shown in Fig. 1:

1) **Contrastive representation learning:** trains a contrastive encoder so that each 8 s PPG window is mapped to a 512-dimensional embedding that is stable across device settings and motion artifacts.
2) **Topology-driven quality discovery:** grounded in the invariance learned in Stage 1, we operate on these encoder-derived representations rather than on the raw waveforms. Specifically, we freeze the encoder, treat each 512-D embedding as a one-dimensional signal, compute a four-scalar PH signature of the embedding landscape and cluster these 4-D signatures with HDBSCAN; the largest dense cluster is labeled clean, while all other points are labeled poor.

The key novelties are (i) the first SSL–TDA fusion for SQA, (ii) cross-device and -sampling rate portability without re-tuning, and (iii) an interpretable four-number signature enabling MCU-level inference.

II. Background

A. Self-supervised learning for physiological signals

Contrastive objectives such as SimCLR [6] and BYOL [7] maximize agreement between two independently augmented views of the same instance; this technique outperforms autoencoders on ECG and PPG [8], [9] and on other biosignals

This work was funded in part by NSF CHEST industrial sponsors.

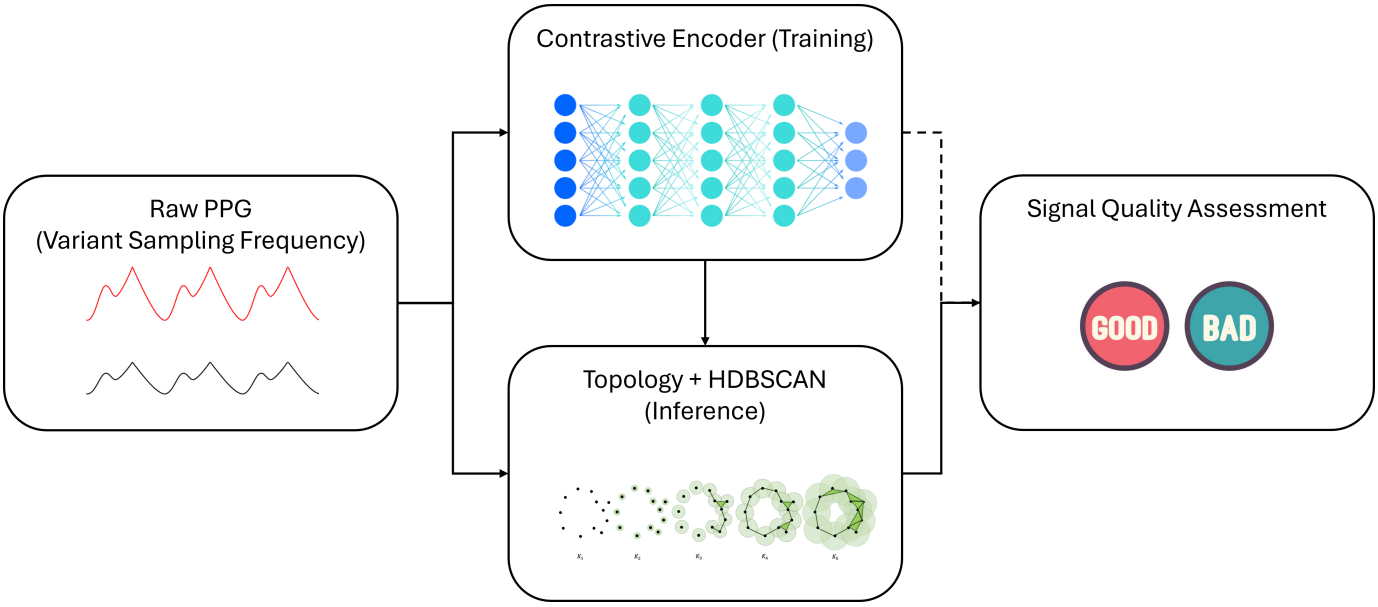

Fig. 1. Proposed two-stage pipeline

by capturing invariance to amplitude scaling and temporal distortion with zero annotation effort.

B. Persistent homology in time-series

TDA quantifies the shape of data. Sublevel-set PH has characterized cardiac periodicity and gait regularity [5], [10]. Clean, quasi-periodic PPG produces long-lived H_1 loops, whereas noisy windows do not, making PH an attractive unsupervised morphology cue. In addition, PH reduces encoder embeddings to morphology-aware scalars, adding an explicit morphological prior—capturing beat regularity versus artifact, and providing a compact and interpretable input to clustering,

C. Density-based clustering for quality discovery

HDBSCAN extends DBSCAN with variable-density cluster extraction and explicit noise labeling [11]. It automatically chooses the number of clusters and handles non-Gaussian shapes—ideal for heterogeneous wrist data where artifacts are rare and scattered.

III. Methodology

A. Corpora and signal conditioning

TABLE I lists the two datasets used in this study.

Why these datasets: WildPPG offers long, mostly clean wrist recordings, whereas We-Be provides lower-rate, motion-rich wrist data. Joint training therefore encourages the encoder to generalize across hardware and noise regimes.

TABLE I
Unlabelled PPG corpora used for pipeline development.

Corpus	Site	Native f_s	Hours	LED
WildPPG [12]	wrist	128 Hz	216	green
We-Be [13], [14]	wrist	25 Hz	60	green

Signal conditioning: A 0.5–8 Hz third-order, zero-phase Butterworth filter removes baseline wander and LED noise. Traces are resampled to a common 25 Hz, z-scored, and segmented into 8 s windows (200 samples, 50 % overlap)

B. Self-supervised representation learning (Contrastive Learning)

a) Loss function: The NT-Xent loss encourages invariance to amplitude and phase jitter—precisely the nuisance factors in wrist PPG—while requiring no annotations.

b) Encoder: A 1-D ResNet-18 processes 1×200 inputs, followed by a projection MLP (512→512→512). The output is ℓ_2-normalized with $\varepsilon = 10^{-6}$.

c) Augmentation strategy: Each view applies a deterministic band-pass filter, then draws two to four of the following transforms:

- **Random crop (keep 50–70%)**: packet loss, strap adjustment, transient motion gaps.
- **Time-warp (±3%)**: natural heart-rate variability, slow sensor drift.
- **Jitter / Gaussian noise (1% SD)**: sensor electronic noise, ambient light flicker.
- **Magnitude scaling (±5%)**: LED drive-current fluctuations, skin perfusion changes.
- **Frequency dropout (narrowband removal)**: ambient light interference, missing harmonics.
- **Circular shift (±1 s) and polarity inversion**: strap orientation errors, polarity mismatches.
- **Segment blackout (10–40 samples)**: short motion spikes (e.g., hand taps).

Empirical studies confirm that augmentations including jitter (Gaussian noise), scaling, time-warp, and polarity inversion reliably mimic motion, noise, and perfusion artifacts in contrastive learning for ECG/PPG signals [9], [15].

d) Training details: By exploring hyperparameter tuning, we use the following: NT-Xent with $\tau = 0.1$; batch size 512; AdamW (learning rate 2×10^{-4}, weight decay 10^{-4}); 200 epochs; mixed precision.

C. Topological signature

We convert each 8 s window into a 512-D embedding using the trained encoder (frozen). Interpreting this embedding as a one-dimensional scalar signal, we compute persistent homology on a 1-D cubical complex (GUDHI) and retain four interpretable features:

$$\left[n_{H_1},\ \Sigma H_1,\ \max H_0,\ \text{mean}\, H_0\right] \in \mathbb{R}^4.$$

These four values summarize the structure in the embedding and are the inputs to HDBSCAN in Stage 2.

D. Unsupervised quality discovery

The 4-D persistence vectors are clustered with HDBSCAN. It adapts the number of clusters automatically, flags sparse points as noise, and handles non-Gaussian shapes—desirable for heterogeneous wrist PPG.

A binary SQI is assigned such that the largest non-noise cluster is deemed clean, and the remaining points (noise plus smaller clusters) are poor.

E. Overall Performance

Because the pipeline is label-free and device-agnostic by design, we evaluate structure quality using standard clustering validity scores (silhouette, Davies–Bouldin, Calinski–Harabasz) rather than supervised accuracy. These metrics capture separability and compactness of the discovered quality strata, which is appropriate when the objective is scalable, cross-device gating without annotation.

IV. Evaluation

A. Encoder convergence

The NT-Xent loss decreases smoothly from $3.44 \rightarrow 0.95$ across 200 epochs, while the mean cosine similarity $(\overline{\cos})$ between the two augmented views rises from $0.67 \rightarrow 0.77$. The coupled evolution of loss and cosine confirms that the encoder learns discriminative directions rather than collapsing to a trivial representation.

B. Quality of topological features

The $150\,000 \times 4$ persistence matrix exhibits a clear morphology gradient: clean windows populate the high-n_{H_1}, high-$\sum H_1$ corner, whereas noisy windows cluster near the origin.

Fig. 2 shows a representative visualization of the clustering outcome: the densest HDBSCAN cluster comprises 76 % of all windows and corresponds to textbook pulsatile traces, explaining the great performance achieved by the SSL–TDA configuration.

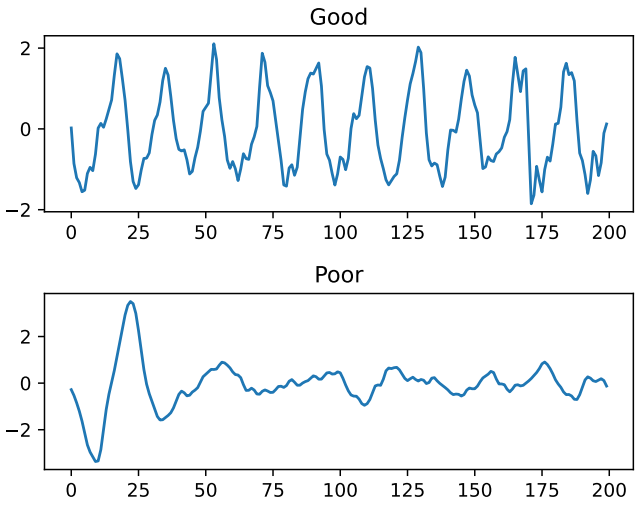

Fig. 2. Visualization of the clustering result.

TABLE II
Ablation on 10 k windows

Configuration	Sil.(↑)	DB(↓)	CH(↑)
SSL+PH+HDBSCAN (SSL-TDA)	**0.72**	**0.34**	6 173
SSL+HDBSCAN (no PH)	0.05	4.60	29
SSL+PH+k-means (no density)	0.39	0.89	**7 177**
SSL+k-means (SSL only)	0.01	7.33	141

C. Ablation study

TABLE II shows the ablation study results. Removing the topological signature (SSL + HDBSCAN) collapses the Silhouette from 0.72 to 0.05: in 512 D, the contrastive embeddings form a diffuse manifold that density-based clustering labels as almost homogeneous. Conversely, retaining PH but swapping HDBSCAN for k-means halves the Silhouette, showing that the density prior is also essential. The full SSL–TDA fusion therefore yields the most compact and well-separated clusters.

D. Unlabeled comparison with existing works

We assess convergent validity by comparing our SSL-TDA pipeline with two baselines on the same unlabeled corpus. For NeuroKit2 [16], we compute per-beat PPG template-matching scores and aggregate them to per-window values (median); for pyPPG [17], we use its 0-1 template-matching SQI. To obtain binary outputs without labels, we prevalence-match thresholds so each method accepts the same fraction $q = 0.24$ of windows (equal to our acceptance rate). On $N = 3600$ windows, it agreed with NeuroKit2 on 82.22% and with pyPPG on 87.44%. These agreements indicate that our label-free method aligns with existing methods on most windows.

V. Discussion

a) Dominant-cluster heuristic: Our current implementation assumes that the largest and densest cluster discovered

by HDBSCAN corresponds to physiologically clean PPG, while smaller or scattered clusters correspond to artifacts. This assumption holds in our corpora, where most windows contain usable signal, but it may break down in regimes dominated by noise. In such cases, the "clean = largest cluster" rule could invert. To mitigate this, one can (i) compute density ratios between the top two clusters and reject segments when the ratio falls below a threshold, (ii) weight clusters by intra-cluster persistence rather than point count, or (iii) use Bayesian nonparametric mixtures that relax the largest-cluster assumption. We note that the clustering framework can accommodate them without retraining the encoder.

b) Practical utility and downstream effects: A natural question is whether the proposed SQA improves downstream analytics such as heart-rate estimation, rhythm classification, or biometric authentication. While we do not include full downstream validation here, prior studies have established that discarding poor-quality PPG segments reduces error rates in heart-rate monitoring and arrhythmia detection, and improves biometric authentication accuracy [4], [18], [19]. Our binary SQI removes roughly 24% of windows in the We-Be dataset; in practice, this would filter the inputs to cardiometric pipelines so that algorithms operate on cleaner segments, reducing spurious beats and missed intervals. We position this work as a modular "quality gate" that can be inserted before such pipelines. A systematic evaluation of downstream benefits, such as pre/post SQI studies on shared labeled corpora—heart rate, rhythm classification, and biometrics, is an important direction for future work.

c) Beyond binary quality: In this paper we report a binary SQI for clarity, assigning the largest dense cluster as clean and all others as poor. However, the clustering framework naturally produces multiple clusters and outlier scores, which could be mapped to finer-grained categories (e.g., clean / borderline / poor) or even a continuous quality index based on cluster density or silhouette distance. Such multi-level outputs may better match downstream applications (e.g., arrhythmia screening, where "borderline" segments should be flagged but not discarded).

d) Multi-modality: Fusing accelerometer and PPG embeddings during contrastive pre-training may boost robustness to motion spikes that currently leak into the clean cluster. In addition, We-Be's LED channels other than green were not exploited; multi-channel PH may further improve robustness. Finally, clinical validation against simultaneous ECG or invasive pressure would solidify the findings.

VI. Conclusion

We presented the first fully unsupervised two-stage pipeline that converts raw wrist-PPG into a binary SQI without device-specific thresholds or expert labels. Rather than optimizing supervised SQA accuracy, we prioritize scalability (no labels), portability (no device-specific re-tuning across 25–128 Hz), and interpretability (four-scalar signature), positioning the method as a practical quality gate for diverse PPG devices, achieving Silhouette 0.72, Davies–Bouldin 0.34, and Calinski–Harabasz 6,173 on 276 h of heterogeneous data. Since it requires zero labels and no hardware calibration, the SSL–TDA framework can serve as a drop-in quality gate for any wrist-based PPG pipeline—paving the way for more reliable heart-rate, rhythm, and biometric-security analytics across the billions of wearables already in use.

References

[1] M. Ragosta and J. L. Kennedy, "Normal waveforms, artifacts, and pitfalls," *Textbook of Clinical Hemodynamics*, pp. 17–55, 2017.

[2] J. Price and T. Goble, "Signals and noise," in *Telecommunications Engineer's Reference Book*. Elsevier, 1993, pp. 10–1.

[3] S. Wabnitz, B. J. Eggleton *et al.*, "All-optical signal processing," *Springer Series in Optical Sciences*, vol. 194, 2015.

[4] T. Pereira, K. Gadhoumi, M. Ma, X. Liu, R. Xiao, R. A. Colorado, K. J. Keenan, K. Meisel, and X. Hu, "A supervised approach to robust photoplethysmography quality assessment," *IEEE journal of biomedical and health informatics*, vol. 24, no. 3, pp. 649–657, 2019.

[5] G. Graff, B. Graff, P. Pilarczyk, G. Jabłoński, D. Gasecki, and K. Narkiewicz, "Persistent homology as a new method of the assessment of heart rate variability," *Plos one*, vol. 16, no. 7, p. e0253851, 2021.

[6] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.

[7] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. Guo, M. Gheshlaghi Azar *et al.*, "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21 271–21 284, 2020.

[8] H. Banville, I. Albuquerque, A. Hyvärinen, G. Moffat, D.-A. Engemann, and A. Gramfort, "Self-supervised representation learning from electroencephalography signals," in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2019, pp. 1–6.

[9] C. Ding and C. Wu, "Self-supervised learning for biomedical signal processing: A systematic review on ecg and ppg signals," *medRxiv*, pp. 2024–09, 2024.

[10] J. Lamar-Leon, R. Alonso-Baryolo, E. Garcia-Reyes, and R. Gonzalez-Diaz, "Persistent homology-based gait recognition robust to upper body variations," in *2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016, pp. 1083–1088.

[11] L. McInnes, J. Healy, S. Astels *et al.*, "hdbscan: Hierarchical density based clustering." *J. Open Source Softw.*, vol. 2, no. 11, p. 205, 2017.

[12] M. Meier, B. U. Demirel, and C. Holz, "Wildppg: A real-world ppg dataset of long continuous recordings," *arXiv preprint arXiv:2412.17540*, 2024.

[13] R. Zhang, R. Fang, M. Orooji, and H. Homayoun, "Introducing webe band: an end-to-end platform for continuous health monitoring," in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2024, pp. 1–5.

[14] R. Fang, S. Hang, R. Zhang, C. Fang, S. Rafatirad, C. Hostinar, and H. Homayoun, "Validation of webe band during physical activities," in *2024 IEEE 20th International Conference on Body Sensor Networks (BSN)*, 2024, pp. 1–4.

[15] S. Soltanieh, A. Etemad, and J. Hashemi, "Analysis of augmentations for contrastive ecg representation learning," in *2022 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2022, pp. 1–10.

[16] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. A. Chen, "Neurokit2: A python toolbox for neurophysiological signal processing," *Behavior research methods*, vol. 53, no. 4, pp. 1689–1696, 2021.

[17] M. Á. Goda, P. H. Charlton, and J. A. Behar, "pyppg: a python toolbox for comprehensive photoplethysmography signal analysis," *Physiological Measurement*, vol. 45, no. 4, p. 045001, 2024.

[18] H. Chen, G. Wang, G. Zhang, P. Zhang, and H. Yang, "Clecg: A novel contrastive learning framework for electrocardiogram arrhythmia classification," *IEEE Signal Processing Letters*, vol. 28, pp. 1993–1997, 2021.

[19] W. Shao, Z. Liang, R. Zhang, R. Fang, N. Miao, E. Kourkchi, S. Rafatirad, H. Homayoun, and C. Fang, "Know me by my pulse: Toward practical continuous authentication on wearable devices via wrist-worn ppg," *arXiv preprint arXiv:2508.13690*, 2025.