

Self-Supervised Learning for Reliable Photoplethysmographic Signal Reconstruction Under Dynamic Motion

Ralph Virtanen

Department of Computer Science, Binghamton University, Binghamton, NY, USA.
ralph.v@binghamton.edu

Ajay Cohen

Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO, USA.
ajaycohen89@missouri.edu

Niklas A. Park

Department of Computer Science, George Mason University, Fairfax, VA, USA.
niklas.park@gmu.edu

Abstract

Photoplethysmography has become central to wearable health monitoring, but its clinical utility is constrained by motion-induced signal corruption. Recent developments in self-supervised learning offer a promising route for reconstructing photoplethysmographic waveforms without relying on large amounts of labeled clinical data. This paper presents a systems-level analysis of self-supervised signal reconstruction under dynamic motion. It examines the architectural trade-offs between reconstruction fidelity, computational burden, and deployability across edge and cloud resources. Rather than focusing narrowly on algorithmic performance, the discussion integrates data infrastructure, lifecycle governance, robustness, fairness, and regulatory policy. The analysis highlights how self-supervised pre-training can exploit large unlabeled corpora of ambulatory physiological recordings, while downstream adaptation must still satisfy clinical validation demands. It further addresses concerns related to population heterogeneity, skin tone, sensor placement, motion intensity, and longitudinal drift. We argue that reliable photoplethysmographic reconstruction is not solely a signal processing problem but a socio-technical systems challenge requiring coordinated advances in representation learning, data curation, model auditing, and deployment practice. The paper concludes with forward-looking perspectives on sustainable and equitable integration of self-supervised reconstruction into wearable health infrastructures.

Keywords

photoplethysmography; self-supervised learning; motion artifact; wearable health; signal reconstruction; responsible artificial intelligence; health systems.

1. Introduction

Wearable photoplethysmography has expanded rapidly as a noninvasive means of estimating heart rate, oxygen saturation, and vascular dynamics in ambulatory populations [1]. The optical signal, however, is highly sensitive to motion, pressure fluctuations, and tissue coupling changes, particularly during everyday physical activities [2]. This sensitivity

undermines the reliability of continuous monitoring in real-world settings. Robust reconstruction of the underlying photoplethysmographic waveform has therefore become a central engineering challenge. The transition from controlled benchtop measurements to dynamic mobile health environments has exposed limitations in conventional filtering and supervised data-driven approaches. Although supervised deep learning can improve motion-tolerant estimation, it typically requires substantial labeled reference data that are costly, difficult to obtain, and often unreliable during intense motion. Self-supervised learning has emerged as a compelling alternative because it can extract structure from large volumes of unlabeled physiological data [3], [4], [5]. The foundational idea is to design pretext tasks that force a model to learn useful representations without requiring manual annotations. These representations can then be fine-tuned for reconstruction or health estimation tasks with limited supervision. In related biomedical signal domains, self-supervised contrastive and predictive objectives have demonstrated the ability to learn robust representations from electroencephalographic and cardiac signals [6], [7], [8]. However, translating these successes to photoplethysmographic reconstruction under dynamic motion demands careful consideration of system architecture, data governance, evaluation rigor, and deployment context. This paper examines those system-level dimensions rather than presenting a single algorithmic solution. The emphasis is on structural trade-offs, infrastructure requirements, robustness, fairness, and policy implications that shape whether self-supervised photoplethysmographic reconstruction can become a dependable clinical and public health technology.

2. Motion Artifacts and Limits of Conventional Reconstruction

Photoplethysmography records pulsatile blood volume changes through light absorption or reflection in microvascular tissue. The measured waveform contains slowly varying baseline changes, respiratory fluctuations, and high-frequency motion interference caused by relative displacement between the sensor and skin, muscle contraction, and ambient light changes [9]. During dynamic motion, the artifact spectrum frequently overlaps with the physiological frequency band of interest, which makes linear filtering insufficient. Motion-tolerant heart rate estimation has been addressed through accelerometer-assisted frameworks that model the relationship between acceleration and optical signal corruption [2]. Such approaches improve average heart rate accuracy, but they often sacrifice waveform morphology and beat-to-beat features that are important for clinical interpretation. Conventional adaptive filtering methods can suppress periodic interference, yet their performance depends on accurate reference inputs and on assumptions about artifact stationarity. Machine learning frameworks can learn more flexible mappings from corrupted optical signals and auxiliary motion channels, but supervised models require paired clean references that are difficult to obtain outside laboratory settings [10]. Adaptive notch filtration has also been proposed to effectively recover photoplethysmographic signals during physical activity by selectively attenuating motion-related components while preserving physiological waveform structures [11]. Although these signal-processing advances remain valuable, they are typically tuned to particular motion patterns and may not generalize across changing activity contexts. The absence of clean ground truth during intense motion creates a fundamental data bottleneck. Consequently, self-supervised learning is attractive because it can learn from corrupted or unlabeled signals directly and can exploit temporal, spatial, and sensor-level regularities in large ambulatory datasets.

3. Self-Supervised Learning as a Systems Alternative

Self-supervised learning encompasses generative, contrastive, and predictive pretext tasks that create supervised signals from the structure of the data itself [3], [5], [12]. In wearable sensing, this approach aligns well with the abundance of continuously recorded but unlabeled photoplethysmographic data. Contrastive methods learn representations by pulling augmented views of the same physiological context together while pushing unrelated contexts apart [3], [8]. Predictive methods train models to forecast future signal segments or to recover masked portions of a waveform, thereby forcing the model to capture physiological dynamics. For photoplethysmographic reconstruction, predictive masking is especially relevant because motion corruption can be treated as a natural form of masking, and the model can learn to reconstruct plausible clean waveform segments from corrupted contexts. The learned representations may then be adapted for downstream tasks such as heart rate variability estimation, respiratory rate extraction, or arrhythmia screening. Studies in cardiac signal representation learning have demonstrated that contrastive objectives can produce features that transfer across patients and recording conditions [7]. In electroencephalography, self-supervised pre-training has reduced the need for large labeled clinical datasets while improving robustness to noise [6]. These cross-domain successes suggest that photoplethysmographic reconstruction systems can benefit from similar pre-training strategies. However, the photoplethysmographic domain introduces distinct challenges, including strong motion baseline drift, dynamic light coupling, and the absence of stable clean reference waveforms during ambulatory monitoring. These challenges imply that pre-training objectives must be robust to non-stationary corruption rather than simply learning from static noise patterns. Predictive masking, for example, can be adapted to recover missing or distorted waveform segments by conditioning on adjacent physiological context, but the model must avoid hallucinating clinically implausible pulse morphology. Contrastive objectives may help align representations across accelerometer and optical channels, thereby encoding motion context without requiring explicit artifact labels [12]. Nevertheless, the choice of augmentation and negative sampling strategy strongly influences whether the learned representation preserves subtle vascular features or collapses to coarse heart rate estimation. From a systems perspective, the self-supervised objective should be viewed as one component within a broader reconstruction pipeline that includes preprocessing, signal quality assessment, multimodal fusion, and downstream clinical inference.

Representation learning for photoplethysmographic reconstruction cannot be separated from downstream deployment constraints. A model that performs well on retrospective benchmark data but demands excessive memory or latency may be unusable on battery-constrained wearables. Similarly, a lightweight model that produces smoothed waveforms may improve heart rate estimation but remove beat-to-beat variability that is essential for stress, sleep, and autonomic assessment. Consequently, system designers must balance reconstruction quality, computational complexity, and physiological informativeness. The choice of backbone architecture interacts with these trade-offs. Temporal convolutional networks provide efficient local feature extraction, while transformer architectures can capture long-range dependencies across respiratory and vasomotor rhythms but introduce higher memory requirements. Hybrid designs that combine causal temporal convolutions with attention mechanisms offer a middle path, yet they demand careful hardware-aware optimization. In addition, the pre-training objective itself can be tuned to favor invariant representations under motion perturbation, thereby reducing the burden on downstream supervised fine-tuning.

4. Architectural Trade-offs and Representation Design

The architectural design of a self-supervised reconstruction system must account for multiple timescales present in photoplethysmographic signals. Cardiac pulsations occur at a relatively high frequency, while respiratory modulation, vasomotor fluctuations, and motion artifacts span slower and faster components. A purely local architecture may fail to capture slow baseline changes, while a purely global architecture may oversmooth beat-level details. Multiscale representations that process the signal at several temporal resolutions provide one structural solution, but they increase the number of trainable parameters and complicate deployment. The representation bottleneck also determines how much motion context can be separated from physiological content. If the bottleneck is too narrow, the model may discard subtle morphological features; if too wide, the model may retain artifact patterns and overfit to spurious correlations. This structural trade-off is central to reliable reconstruction under dynamic motion.

A further architectural consideration is whether the reconstruction system should be trained centrally or in a federated manner across devices and clinical sites. Centralized training can exploit large heterogeneous datasets and simplify experimentation, but it raises privacy risks and creates dependence on continuous data transfer. Federated learning offers an alternative by keeping raw physiological data on user devices while sharing model updates [13]. However, federated optimization introduces additional systems challenges, including heterogeneity across hardware, non-identically distributed data, communication cost, and the need for robust aggregation under varying data quality. The selection of a federated or centralized architecture is not merely a technical decision but a governance choice that affects data stewardship, participant privacy, and long-term maintainability. Systems researchers should therefore treat architecture selection as a socio-technical trade-off involving accuracy, equity, and regulatory compliance.

5. Data Infrastructure and Governance

The performance of self-supervised photoplethysmographic reconstruction depends on the breadth and diversity of the unlabeled data used during pre-training. Large-scale wearable deployments can generate continuous multi-day recordings across diverse populations, activities, and environmental conditions. However, data volume alone does not guarantee useful representation learning. The curation of ambulatory photoplethysmographic corpora must address missing segments, sensor disconnection, inconsistent sampling rates, and varying device characteristics. Data infrastructure should support versioning, provenance tracking, and quality labeling without imposing overly restrictive inclusion criteria that reduce population diversity. Moreover, continuous physiological data are highly identifiable and sensitive. Governance frameworks must incorporate informed consent, data minimization, purpose limitation, and auditability. The use of federated or distributed learning approaches can reduce the need to centralize raw waveforms [13], but they do not eliminate the need for clear data governance because model updates can still leak information under certain adversarial conditions.

Another infrastructure challenge is the absence of high-quality clean references during intense motion. In supervised learning, ground truth may be acquired from simultaneous electrocardiogram or motion-controlled laboratory protocols, but these references are limited in scale and ecological validity. Self-supervised learning reduces dependence on such labels, yet evaluation still requires some reference signals for validation and regulatory approval. Therefore, systems must maintain carefully curated validation datasets that cover diverse motion intensities and demographic groups. These datasets should be collected with

standardized protocols while also including naturalistic free-living conditions to test generalization. Data infrastructure should also support continuous monitoring of concept drift after deployment, because sensor hardware, firmware, and population characteristics evolve over time. Without such infrastructure, a reconstruction model may silently degrade in ways that are not detected by periodic performance reports.

6. Robustness and Generalization Across Populations

Robustness to population heterogeneity is a central concern for photoplethysmographic reconstruction systems. Optical signals are influenced by skin tone, epidermal thickness, vascular density, perfusion, and subcutaneous fat. If the pre-training corpus underrepresents darker skin tones, the learned representation may encode spurious relationships between melanin concentration and signal corruption, leading to biased reconstruction quality. Studies of wearable optical sensors have demonstrated that measurement errors can be higher for individuals with darker skin tones [16]. Self-supervised learning does not automatically correct such disparities; it may magnify them if the training distribution is skewed or if the pretext task emphasizes superficial signal statistics. Therefore, fairness must be incorporated into data collection, model evaluation, and post-deployment monitoring. Stratified performance reporting across demographic groups is necessary to detect differential reconstruction fidelity.

Generalization across sensor placement is another underappreciated challenge. A model trained on wrist-worn photoplethysmography may not transfer to finger, chest, or ear locations because the optical path geometry and motion coupling differ. Domain adaptation methods can align representations across sensor placements or device types by learning adversarial or distribution-matching objectives [14]. However, such adaptation requires access to unlabeled data from target domains, which may raise privacy concerns and increase system complexity. Beyond sensor placement, motion intensity varies from sedentary activities to vigorous exercise. A reconstruction system that has only seen moderate motion may fail under high-intensity acceleration. Domain generalization strategies that expose the model to diverse augmentation patterns can improve robustness, but they must not obscure rare physiological events. These trade-offs illustrate that robustness cannot be optimized solely through architectural modifications; it requires a coordinated approach across data infrastructure, evaluation protocols, and algorithmic design.

7. Fairness, Explainability, and Evaluation Governance

Fairness in self-supervised photoplethysmographic reconstruction is not limited to demographic parity. It includes equitable access to reliable monitoring across skin tones, body types, age groups, and health conditions. If a reconstruction system performs well for young healthy adults but poorly for older adults with low perfusion, its clinical utility may be dangerously uneven. Fairness assessment should be embedded in the evaluation pipeline rather than treated as a post hoc checkbox. This requires diverse reference datasets, disaggregated error metrics, and explicit definition of acceptable performance bounds for each subgroup. Work on fairness in medical machine learning has emphasized that algorithmic bias often arises from structural disparities in data collection and label availability [15]. Self-supervised models may mitigate label scarcity but cannot resolve underlying representation imbalances without deliberate governance.

Explainability is another critical dimension. Clinicians are unlikely to rely on reconstructed photoplethysmographic waveforms unless they can understand when and why the model may

produce plausible but incorrect morphology. Interpretable architectures or explanation tools can provide heatmaps over input segments, highlight motion contexts that degrade reconstruction, or quantify uncertainty [18]. Model cards and structured reporting frameworks can communicate intended use, limitations, and subgroup performance to users and regulators [17]. Such transparency mechanisms are essential for building trust in wearable health systems, but they also introduce additional engineering overhead. Systems designers must decide how much explanation is required for different use cases. A consumer wellness application may tolerate lower interpretability than a hospital-grade arrhythmia screening tool. Governance policies should distinguish between these risk levels while ensuring that high-stakes clinical decisions are never made solely on opaque reconstructions without human or rule-based safeguards.

8. Deployment, Sustainability, and Resource Management

Deployment of self-supervised reconstruction models on wearable devices requires careful attention to energy consumption, memory footprint, and real-time constraints. Photoplethysmographic monitoring is continuous, which means that even modest per-sample energy costs accumulate rapidly over days and weeks. Edge intelligence frameworks can offload heavy computation to nearby gateways or cloud services, but this introduces latency and connectivity dependence [19]. A hybrid deployment model may be appropriate, where initial on-device processing handles quality gating and simple artifact suppression while more complex reconstruction occurs on a paired smartphone or remote server when connectivity and battery allow. This hierarchical approach balances responsiveness with computational sustainability, but it complicates system design and raises questions about data transmission.

Sustainability also includes the environmental and operational overhead of training large self-supervised models. Pre-training on massive physiological datasets can consume significant computational resources, especially when using transformer architectures and extensive hyperparameter search. The research community should consider whether performance gains justify the energy footprint and whether smaller, carefully designed models can achieve comparable reliability for reconstruction tasks. Model compression, quantization, and knowledge distillation can reduce inference costs, but they may interact with the self-supervised features in nontrivial ways. Lifecycle management further requires monitoring model drift, updating parameters as populations and hardware change, and maintaining backward compatibility with legacy devices. Without sustainable deployment practices, even highly accurate reconstruction models may be abandoned or fail in the field, undermining the long-term value of self-supervised approaches.

9. Clinical Translation and Multimodal Integration

Clinical translation of self-supervised photoplethysmographic reconstruction requires more than improved waveform similarity. Clinicians and regulators need evidence that reconstructed signals preserve features relevant to diagnosis, such as pulse arrival time, beat morphology, and respiratory modulation. Validation should include reference-standard comparisons under controlled motion protocols and prospective free-living studies. Photoplethysmography-based atrial fibrillation detection and other cardiovascular screening applications depend on reliable beat-to-beat features [20]. If reconstruction smooths or distorts these features, downstream algorithms may lose sensitivity for clinically meaningful events. Therefore, clinical validation must be task-specific rather than relying solely on generic reconstruction error.

Multimodal integration offers a promising path toward more reliable motion-aware inference. Combining photoplethysmography with accelerometry, electrocardiography, or contextual activity data can provide complementary information that assists in separating physiological signals from motion artifacts. Foundation models trained across multiple biosignals may further improve robustness and transfer learning [21], but they also raise new governance questions about data harmonization and cross-modal leakage. The integration of additional sensors increases device cost and complexity, which may limit access in resource-constrained settings. System designers should assess whether multimodal inputs are necessary for the target application or whether robust self-supervised photoplethysmographic reconstruction alone can meet clinical needs. This assessment should include not only accuracy but also usability, equity, and regulatory burden. Furthermore, data protection regulations such as the GDPR impose specific obligations on processing health-related data and on automated decision-making, requiring a clear legal basis, purpose specification, and data subject rights [22]. These regulatory requirements shape how self-supervised models can be trained, evaluated, and updated in real-world health systems.

10. Conclusion

Self-supervised learning holds substantial promise for reliable photoplethysmographic signal reconstruction under dynamic motion, but its success depends on system-level considerations beyond algorithmic novelty. The choice of pre-training objectives, architecture, and data infrastructure shapes whether learned representations preserve clinically relevant morphology. Robustness across populations, sensor placements, and motion intensities must be explicitly engineered and validated. Fairness, explainability, and governance frameworks are needed to prevent biased performance and to support responsible deployment. Sustainable edge and cloud strategies are essential for continuous wearable monitoring. Clinical translation requires task-specific validation and careful integration with multimodal signals. The required reference on adaptive notch filtration illustrates how targeted signal-processing techniques can complement data-driven reconstruction approaches [11]. Future work should focus on building transparent, federated, and hardware-aware self-supervised systems that are evaluated not only by reconstruction error but also by equity, robustness, and long-term real-world reliability.

References

1. Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), R1–R39. <https://doi.org/10.1088/0967-3334/28/3/R01>
2. Zhang, Z., Pi, Z., & Liu, B. (2015). TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on Biomedical Engineering*, 62(2), 522–531. <https://doi.org/10.1109/TBME.2014.2359372>
3. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. *Proceedings of the 37th International Conference on Machine Learning*, 1597–1607.
4. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16000–16009.

5. Jaiswal, A., Babu, A. R., Zadeh, M. Z., Banerjee, D., & Makedon, F. (2021). A survey on contrastive self-supervised learning. *Technologies*, 9(1), 2. <https://doi.org/10.3390/technologies9010002>
6. Banville, H., Chehab, O., Hyvärinen, A., Engemann, D.-A., & Gramfort, A. (2021). Uncovering the structure of clinical EEG signals with self-supervised learning. *Journal of Neural Engineering*, 18(4), 046020. <https://doi.org/10.1088/1741-2552/abca18>
7. Kiyasseh, D., Zhu, T., & Clifton, D. A. (2021). CLOCS: Contrastive learning of cardiac signals across space, time, and patients. *Proceedings of the 38th International Conference on Machine Learning*, 5606–5615.
8. Mehari, T., & Strodthoff, N. (2022). Self-supervised representation learning from 12-lead ECG data. *Computers in Biology and Medicine*, 141, 105114. <https://doi.org/10.1016/j.compbiomed.2021.105114>
9. Wijshoff, R. W. C. G. R., Mischi, M., & Aarts, R. M. (2012). Reducing motion artifacts in photoplethysmograms by using relative sensor motion: phantom study. *Journal of Biomedical Optics*, 17(11), 117007. <https://doi.org/10.1117/1.JBO.17.11.117007>
10. Charlton, P. H., Kyriacou, P. A., Mant, J., & Marozas, V. (2022). Wearable photoplethysmography for cardiovascular monitoring. *Proceedings of the IEEE*, 110(3), 355–381. <https://doi.org/10.1109/JPROC.2022.3149785>
11. Zheng, X., Dwyer, V. M., Barrett, L. A., Derakhshani, M., & Hu, S. (2022). Adaptive notch-filtration to effectively recover photoplethysmographic signals during physical activity. *Biomedical Signal Processing and Control*, 72, 103303.
12. Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J. (2021). Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 35(1), 857–876. <https://doi.org/10.1109/TKDE.2021.3090866>
13. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3, 119. <https://doi.org/10.1038/s41746-020-00323-1>
14. Ganin, Y., & Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. *Proceedings of the 32nd International Conference on Machine Learning*, 1180–1189.
15. Chen, I. Y., Szolovits, P., & Ghassemi, M. (2021). Can AI help reduce disparities in general medical and mental health care? *AMA Journal of Ethics*, 23(2), E167–E179. <https://doi.org/10.1001/amajethics.2021.167>
16. Colvonen, P. J., DeYoung, P. N., Bosompra, N. O. A., & Owens, R. L. (2020). Limiting racial disparities and bias in wearable devices. *Sleep*, 43(9), zsaa159. <https://doi.org/10.1093/sleep/zsaa159>
17. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. <https://doi.org/10.1145/3287560.3287596>

18. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
19. Zhou, Z., Chen, X., Li, E., Zeng, L., Luo, K., & Zhang, J. (2019). Edge intelligence: Paving the last mile of artificial intelligence with edge computing. *Proceedings of the IEEE*, 107(8), 1738–1762. <https://doi.org/10.1109/JPROC.2019.2918951>
20. Pereira, T., Tran, N., Gadhoumi, K., Pelter, M. M., Do, D. H., Lee, R. J., Colorado, R., Meisel, K., & Hu, X. (2020). Photoplethysmography based atrial fibrillation detection: a review. *npj Digital Medicine*, 3, 3. <https://doi.org/10.1038/s41746-019-0207-9>
21. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. *Nature*, 616(7956), 259–265. <https://doi.org/10.1038/s41586-023-05881-4>
22. Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A practical guide*. Springer.