

Deep Learning-Based Motion Artifact Removal for Wearable Photoplethysmographic Signal Enhancement

Martin Patiers

Department of Computer Science, Binghamton University, Binghamton, NY, USA.
peters1973@binghamton.edu

Jordan M. Bryant

Department of Computer Science, University of New Hampshire, Durham, NH, USA.
jordan.work@unh.edu

Grther Teylar

Department of Computer Science, University of North Texas, Denton, TX, USA.
ataylor@unt.edu

Abstract

Wearable photoplethysmography has become a central sensing modality for continuous cardiovascular monitoring, yet its practical value is constrained by motion artifacts that corrupt the underlying blood volume pulse waveform. Deep learning-based artifact removal has emerged as a promising strategy because it can learn complex mappings between corrupted and clean representations without relying solely on hand-engineered assumptions. This paper provides a system-level analysis of deep learning-based motion artifact removal for wearable photoplethysmographic signal enhancement. It examines architectural choices, training data infrastructure, model governance, deployment models, robustness, fairness, and sustainability. The discussion treats artifact removal as a sociotechnical infrastructure rather than an isolated signal processing task. It explores how convolutional, recurrent, attention-based, and encoder-decoder designs interact with the energy, memory, and latency constraints of wearable platforms. The paper also considers the governance implications of data provenance, annotation uncertainty, performance drift, and algorithmic bias. In particular, it addresses how differential optical behavior across skin pigmentation, sensor placement, and motion conditions can produce uneven performance if not explicitly managed during data collection and validation. The work argues that safe deployment requires integrating deep learning models with continuous monitoring, edge-based inference, and regulatory oversight. It further highlights the importance of multimodal sensor fusion, uncertainty estimation, and federated learning as pathways toward robust and equitable wearable cardiovascular sensing. The conclusion synthesizes these perspectives and identifies future research directions that connect signal enhancement, infrastructure governance, and clinical accountability.

Keywords

photoplethysmography; motion artifact removal; deep learning; wearable sensors; edge computing; algorithmic fairness; digital health governance.

1. Introduction

Photoplethysmography provides a noninvasive optical measurement of blood volume changes in the microvascular bed and has become a foundational sensing technique for wearable health monitoring [1]. The proliferation of wrist-worn, finger-worn, and earbud devices has dramatically increased the availability of continuous photoplethysmographic data outside controlled clinical environments [2]. However, the photoplethysmographic signal is highly sensitive to sensor displacement, tissue compression, local perfusion changes, and ambient light interference. These influences generate motion artifacts that obscure pulse morphology and degrade the accuracy of derived parameters such as heart rate, heart rate variability, oxygen saturation, and blood pressure estimates [3]. Traditional signal processing approaches, including bandpass filtering, adaptive filtering, and signal decomposition, can attenuate some artifacts but often fail when the noise spectrum overlaps with the physiological frequency range [4]. Deep learning-based removal therefore represents a structural departure from fixed signal processing pipelines toward learned representations that can capture highly nonlinear artifact dynamics. This shift introduces a new set of systems challenges related to model architecture, data governance, computational deployment, fairness, and sustainability. Early work on deep photoplethysmography has demonstrated that convolutional neural networks can estimate heart rate during intense activity, suggesting that learned models can outperform conventional methods under certain conditions [5].

The purpose of this paper is not to propose a single network topology but to examine the broader system within which deep learning-based artifact removal operates. A robust wearable monitoring infrastructure must coordinate the sensor front end, the inference engine, the data management layer, the clinical validation process, and the policy environment. These components interact in ways that are not captured by reconstruction error alone. For example, a model with excellent artifact removal in laboratory settings may fail in field deployments because of sensor shift, population heterogeneity, or battery constraints. Similarly, a highly accurate cloud-based model may introduce privacy risks that are unacceptable in real clinical workflows. These interactions justify an interdisciplinary perspective that links signal processing with systems engineering, data governance, and regulatory science.

The remainder of this paper is organized as follows. Section 2 reviews relevant background and related work. Section 3 examines system architecture trade-offs. Section 4 addresses data governance and signal quality infrastructure. Section 5 considers model design and robustness. Section 6 discusses deployment and edge computing. Section 7 integrates fairness, clinical validation, and policy. Section 8 analyzes sustainability and lifecycle concerns. Section 9 outlines future directions, and Section 10 concludes.

2. Background and Related Work

Wearable photoplethysmography has been extensively reviewed as a cardiovascular monitoring tool, with attention to waveform morphology, sensor placement, tissue optics, and clinical applicability [6]. In parallel, model-based and machine-learning methods have been developed for heart rate estimation during physical exercise, where motion artifact corruption is most severe [7]. Sensor fusion approaches that combine photoplethysmography with accelerometer signals have improved motion cancellation, but they demand precise synchronization, sensor calibration, and robust spectral peak selection [8]. Time-varying spectral filtering strategies have also been proposed to reconstruct motion-corrupted heart rate signals during intense physical activity, demonstrating that adaptive signal models can improve reliability in challenging conditions [9]. Adaptive filtering investigations have further explored notch-filtration strategies for recovering photoplethysmographic signals

during physical activity, and such approaches illustrate the continuing value of combining reference motion information with adaptive suppression mechanisms [10]. Multiwavelength sensing has been studied as a complementary route to motion artifact reduction because different optical wavelengths respond differently to tissue deformation, blood flow, and pigmentation [11]. Clinical deployments in atrial fibrillation populations have shown that wrist-based photoplethysmography can track heart rate and inter-beat intervals, but motion and rhythm irregularity together create complex validation challenges [12].

These studies collectively indicate that artifact removal is not exclusively a signal processing problem. It is shaped by sensor design, reference signal availability, patient characteristics, activity context, and the clinical use case. Conventional approaches generally rely on assumptions about the stationarity or spectral separability of artifact and signal. Deep learning relaxes some of these assumptions but introduces data dependence, model complexity, and deployment uncertainty. The transition from hand-engineered filters to learned denoisers therefore changes the location of risk within the system. Instead of validating a fixed filter bank, engineers must validate a learned mapping whose behavior may vary across populations, environments, and hardware revisions.

In this context, artifact removal should be conceptualized as a critical preprocessing stage within a larger analytical pipeline. The output of the deep learning model is rarely the final clinical endpoint. It feeds into peak detectors, beat classification models, arrhythmia detectors, and physiological parameter estimators. Errors at the artifact removal stage propagate downstream and can be amplified by subsequent processing. Therefore, the design of artifact removal models must account for downstream requirements such as waveform morphology preservation, inter-beat interval accuracy, and the ability to flag unreliable segments. This systems view is essential for understanding why reconstruction error alone is an incomplete measure of model performance.

3. System Architecture for Deep Learning-Based Artifact Removal

A deep learning-based artifact removal system can be decomposed into several architectural layers: the optical sensing front end, synchronized motion references, signal preprocessing, the neural inference engine, postprocessing, and downstream physiological analytics. The central architectural decision is whether the model should reconstruct the full photoplethysmographic waveform or instead estimate only a derived parameter such as heart rate. Waveform-domain reconstruction preserves clinically meaningful morphology, including the systolic upstroke and diastolic foot, but it imposes high output dimensionality and can be computationally expensive. Parameter-domain models are typically lighter and can be deployed more easily on constrained hardware, but they discard waveform features that may be needed for advanced diagnostics. This trade-off between clinical richness and computational feasibility is fundamental to wearable system design.

Convolutional encoder-decoder architectures are widely used for waveform reconstruction because they can learn local artifact features and compress the signal into a latent representation before generating a clean output. These models align well with the hierarchical structure of photoplethysmographic signals and can be implemented with relatively low memory overhead. Recurrent networks offer strong temporal modeling capabilities for sequential physiological data, but they may be harder to optimize, and their sequential dependencies can complicate parallel execution. Attention mechanisms provide long-range dependency modeling and can improve performance under irregular artifact patterns, yet their memory and computation requirements often exceed the resources of low-power wearable

processors. System architects must therefore select a model family based on the intended deployment environment, the required output resolution, and the tolerance for inference latency.

The availability of auxiliary sensor channels also shapes the architecture. Many wearable devices include accelerometers or gyroscopes that capture motion references. These signals can be concatenated with the photoplethysmographic inputs, passed through separate encoders, or used as conditioning signals to inform the artifact removal process. Multiwavelength photoplethysmography further expands the input space and can improve robustness because different wavelengths carry complementary information about tissue motion and blood flow [11]. However, each additional sensor channel increases power consumption, data synchronization complexity, and the risk of sensor failure. A well-governed system must decide whether the marginal robustness gain justifies these operational costs.

Another architectural consideration concerns uncertainty estimation. Deterministic models may produce plausible-looking waveforms even when the input is too corrupted to support reliable reconstruction. Uncertainty-aware models can flag such segments and prevent downstream algorithms from making unsafe decisions. This ability is particularly important for ambulatory monitoring, where no clinical expert is present to inspect the waveform. Uncertainty estimation can be implemented through probabilistic layers, ensemble methods, or auxiliary prediction heads. Although these mechanisms increase computational load, they improve system safety and support regulatory approval by making model limitations more transparent. Thus, architecture selection is not merely an engineering optimization; it is also a governance decision that reflects the system's failure logic and clinical accountability.

4. Data Governance and Signal Quality Infrastructure

Deep learning models for healthcare applications depend on data pipelines that are more complex and more consequential than those used in many consumer machine learning tasks [13]. In medicine, data distribution, annotation accuracy, and population representativeness directly influence patient safety [14]. When training data encode structural inequalities or measurement biases, the resulting model may perform unevenly across groups [15]. Ethical analyses of algorithmic decision making further emphasize the need for transparency, responsibility, and contestability in systems that affect human welfare [16]. For wearable photoplethysmographic artifact removal, these concerns are especially salient because training data are often collected from convenient populations and under laboratory conditions that do not represent real-world variability.

Signal quality infrastructure must therefore begin with rigorous data acquisition protocols. Motion artifact corruption should be systematically induced across different activity types, sensor placements, and environmental conditions. Ground truth physiological references should be collected using electrocardiography or other clinically validated devices. The annotation process must account for timing offsets, heart rate variability, ectopic beats, and periods during which the reference itself is unreliable. Without this level of rigor, models may learn to remove artifacts in ways that also remove physiological information or that fail outside the training distribution.

Data governance also includes provenance tracking, consent management, and interoperability. Photoplethysmographic data can contain sensitive information about cardiac function, stress, and physical activity. Systems must ensure that data are collected with

informed consent and that reuse is consistent with stated purposes. Federated learning and distributed data processing can reduce the need to centralize raw waveforms, but they introduce new governance challenges related to model aggregation, auditability, and cross-institutional data drift. The governance framework must therefore extend beyond privacy compliance to include technical mechanisms for monitoring data quality, population coverage, and model behavior over time.

A robust signal quality infrastructure should treat data curation as a continuous process rather than a one-time training exercise. As wearable devices evolve, as sensors age, and as user populations diversify, the distribution of valid signals and artifacts changes. Continuous data collection from deployed devices, with appropriate consent and privacy protections, enables the detection of performance drift and the construction of new training cohorts. However, this approach must be balanced against the risk of reinforcing deployment biases. If models are retrained primarily on data from users who can afford newer devices or who are more active, the system may become increasingly tuned to those subgroups. Governance mechanisms must therefore include population-level auditing and explicit rebalancing of training data.

5. Model Design Trade-Offs and Robustness

The design of deep learning models for motion artifact removal involves multiple trade-offs that cannot be resolved by reconstruction accuracy alone. A model may be highly expressive and capable of fitting complex training data, yet this expressiveness can also lead to overfitting and poor generalization when sensor conditions change. Regularization, data augmentation, and architectural constraints are therefore necessary to ensure that the model learns physiologically meaningful representations rather than spurious correlations. The balance between capacity and generalization is particularly delicate in wearable photoplethysmography because artifact distributions are highly variable across users and activities.

Supervised learning has been the dominant paradigm in many photoplethysmography studies, with paired corrupted and clean signals used to train denoising models [5]. However, clean references are often difficult to obtain in ambulatory settings. Laboratory motion protocols may fail to capture the richness of everyday movement, while ambulatory data may lack reliable ground truth. Self-supervised and semi-supervised approaches offer a potential solution by learning from unlabeled photoplethysmographic data, but they require careful design to avoid learning representations that discard diagnostically important signal components. The choice of training paradigm is therefore a systems decision that affects data requirements, annotation costs, and model robustness.

Domain adaptation is another important consideration. A model trained on one sensor type, placement, or population may not transfer to another. This is especially relevant given the wide variety of wearable platforms, each with different optical components, sampling rates, and signal conditioning circuits. Domain-invariant representations can improve robustness across sensor variations, but they may also reduce sensitivity to clinically relevant differences in waveform morphology. System designers must decide whether to prioritize a single general model, a family of device-specific models, or adaptive fine-tuning at deployment. Each option has implications for maintenance, regulatory approval, and user experience.

The interaction between artifact removal and downstream physiological estimation deserves particular attention. Some artifact removal models are optimized to improve heart rate estimation accuracy, but they may distort beat-to-beat intervals or introduce artificial

waveform features. If the same signal is later used for oxygen saturation estimation or arrhythmia detection, these distortions can be harmful. Robust design therefore requires multiobjective evaluation that considers waveform fidelity, inter-beat interval accuracy, and downstream clinical performance. It also requires collaboration between signal processing engineers, clinical researchers, and regulatory experts, because the meaning of robustness differs across these communities.

Cross-domain comparisons can help situate photoplethysmographic artifact removal within the broader landscape of biomedical signal denoising. Electrocardiographic and electroencephalographic denoising have benefited from decades of standardized datasets and benchmark frameworks. Photoplethysmography has fewer widely accepted benchmark resources, partly because of the diversity of sensor platforms and motion conditions. This lack of standardization makes it difficult to compare models fairly and to build a cumulative evidence base. The field would benefit from shared datasets, common evaluation protocols, and transparent reporting of model limitations. Such infrastructure is not merely a convenience for researchers; it is essential for regulatory confidence and clinical translation.

6. Deployment and Edge Computing Considerations

Deployment decisions determine whether a deep learning artifact removal model can operate safely and efficiently in real-world wearable environments. Edge intelligence research has shown that distributing computation across sensor nodes, smartphones, gateways, and cloud servers can improve responsiveness and privacy while reducing dependence on continuous connectivity [17]. In wearable photoplethysmography, on-device inference is attractive because it can process signals locally, minimize data transmission, and provide immediate feedback to users. However, edge deployment also imposes strict constraints on model size, memory bandwidth, and energy consumption. These constraints often require model compression, quantization, pruning, or the use of specialized neural processors.

The practical implementation of artificial intelligence technologies in medicine highlights the importance of workflow integration, validation, and post-deployment monitoring [18]. A photoplethysmographic artifact removal model that runs on a microcontroller may be energy efficient, but if it cannot be updated easily or if its performance degrades silently, the clinical system may fail. Hybrid architectures that perform lightweight denoising on the device and more complex analysis in the cloud can balance latency and analytical depth, but they introduce data transfer and synchronization risks. The boundary between edge and cloud must be designed with careful attention to privacy, consent, and data minimization. For example, a system might transmit only reconstructed waveforms or derived features rather than raw photoplethysmographic data, yet this can reduce the ability to audit model decisions and investigate errors.

Energy consumption is a particularly important constraint for wearable devices that are expected to operate continuously for days or weeks. Deep learning inference can be a significant source of battery drain if not optimized. Architectural choices such as depth, channel width, and temporal resolution directly affect energy use. Low-power modes may require the model to operate intermittently, which can create gaps in monitoring and complicate artifact removal across transitions between activity states. The deployment architecture should therefore include power-aware scheduling and adaptive inference strategies. A model might use a lightweight denoiser during low-motion periods and activate a more complex model only when motion artifacts exceed a threshold. Such approaches

require reliable artifact detection and a policy for switching models without introducing discontinuities.

The regulatory status of deployed models also affects deployment choices. In many jurisdictions, software that processes physiological signals for medical purposes may be regulated as a medical device. Updates to model weights, architecture changes, and retraining procedures may require regulatory review. A cloud-based model can be updated centrally, but this may create versioning and traceability challenges. An on-device model may be more difficult to update, but it can offer a stable and auditable configuration. Deployment planning should therefore be coordinated with quality management systems, risk management processes, and regulatory affairs teams. The engineering goal of rapid model improvement must be balanced against the governance goal of maintaining a controlled and transparent system.

7. Fairness, Clinical Validation, and Policy Implications

Fairness in wearable photoplethysmography is not an abstract concern. The World Health Organization has emphasized that artificial intelligence for health should be designed and governed to promote equity, safety, and public trust [19]. A particularly relevant example is the documented racial bias in pulse oximetry measurement, where skin pigmentation can affect the accuracy of optical oxygen saturation estimates [20]. Photoplethysmography shares similar optical principles, and sources of inaccuracy include skin tone, tissue composition, sensor contact pressure, and ambient light [21]. If deep learning artifact removal models are trained on populations with limited diversity, they may fail to generalize to users with darker skin tones or different tissue optics. This failure is not simply a matter of lower average accuracy; it can result in systematic disparities that have clinical consequences.

Clinical validation studies of wearable photoplethysmography, such as passive detection of atrial fibrillation using commercial smartwatches, show that large-scale ambulatory monitoring can provide useful information, but they also reveal the complexity of validating algorithms across heterogeneous populations [22]. Validation must go beyond aggregate metrics and examine performance across subgroups defined by age, sex, skin pigmentation, body mass index, cardiovascular status, and activity patterns. Subgroup analysis requires sufficiently large and representative datasets, which are costly to assemble. Nevertheless, this is necessary to ensure that the benefits of motion artifact removal are distributed fairly and that the risks are not concentrated in vulnerable populations.

Policy implications extend to the governance of training data, the transparency of model limitations, and the accountability of manufacturers. A deep learning artifact removal model may be embedded within a commercial wearable device, but users and clinicians may not know when the model is uncertain or when its performance is degraded. Disclosure mechanisms such as quality indicators, data adequacy labels, and uncertainty flags can help users interpret outputs. Regulatory frameworks should require post-market surveillance that explicitly monitors performance across demographic and environmental strata. This is a departure from traditional device regulation, which often emphasizes premarket testing under standardized conditions.

Fairness and clinical validation are also tied to the construction of evaluation benchmarks that reflect the intended deployment population rather than a convenient laboratory cohort. Benchmarks that rely heavily on young, healthy subjects performing standardized treadmill protocols may yield spuriously optimistic results

when models are deployed among older adults, individuals with cardiovascular disease, or users with darker skin pigmentation. A system-level approach therefore requires validation across multiple dimensions, including demographic coverage, motion intensity, sensor placement, and environmental illumination. Such validation is expensive, but it is necessary to ensure that the infrastructure does not silently encode uneven performance. In addition to demographic fairness, temporal fairness is an often-overlooked dimension. Wearable models may perform differently during different times of day, seasons, or activity transitions because optical coupling, perfusion, and motion patterns vary across these contexts. A model validated during daytime walking may fail during sleep, intense exercise, or periods of rapid posture change. Validation protocols should include longitudinal ambulatory data rather than only short laboratory sessions. Furthermore, the output of a motion artifact removal model is not a neutral waveform; it is a filtered representation that conditions subsequent diagnostic claims. Therefore, fairness cannot be reduced to a single error metric but must be evaluated through the entire analytical chain, from raw sensor input to final clinical output.

From a policy perspective, manufacturers and developers must be able to trace model decisions, document training data limitations, and respond to reports of degraded performance in specific subgroups. This requires a quality management system that treats machine learning models as living components rather than static software. Post-market surveillance should include algorithmic audits that examine performance across demographic and environmental strata, and these audits should be conducted at intervals that reflect real-world deployment drift. Without such governance, deep learning-based artifact removal may repeat the historical failures of optical devices that did not account for skin pigmentation. In contrast, a governance framework that integrates technical validation with equity monitoring can support trustworthy digital health infrastructure. The regulatory challenge is that algorithms are updated more frequently than traditional hardware, and each update may alter the risk profile in ways that are not visible to clinicians or patients. Transparency mechanisms, including quality indicators and uncertainty labels, can help users interpret outputs and make informed decisions. Such mechanisms should be designed with input from patient advocates, clinicians, and regulatory bodies, because the meaning of a waveform quality score may differ across clinical contexts.

8. Sustainability and Lifecycle Considerations

Sustainability considerations for wearable deep learning systems include more than battery life, although energy efficiency remains a central constraint. Training large denoising models can consume significant computational resources, particularly when extensive hyperparameter search and repeated data augmentation are required. The environmental cost of model training is increasingly recognized as a governance issue, and it should be weighed against the modest improvements in artifact removal that a larger model may provide. This does not mean that model capacity should always be minimized, but rather that developers should adopt efficiency-aware design principles such as knowledge distillation, structured pruning, and low-rank decomposition. Such approaches can reduce training and inference costs while preserving clinically relevant signal fidelity. At the edge, energy consumption must be considered jointly with battery capacity and device longevity. Frequent deep learning inference can accelerate battery degradation and reduce the usable life of the wearable device. This creates a lifecycle trade-off: a model that provides continuous high-fidelity denoising may also shorten the hardware replacement cycle, with implications for electronic waste and cost. Power-aware scheduling, adaptive sampling, and duty cycling can mitigate these effects,

but they require coordination between the artifact removal model and the broader system resource manager. A sustainable deployment is therefore one that treats computation as a finite resource and allocates it according to clinical need, not merely according to algorithmic convenience.

Model update lifecycles also raise sustainability questions. A model that is retrained frequently may incorporate new data and improve performance, but frequent updates can create version fragmentation and require additional validation. This is particularly challenging in regulated medical device contexts, where each model version may need to be linked to a specific validation record. A modular approach that updates only the artifact removal front end while leaving downstream clinical algorithms unchanged may reduce the regulatory burden, but it still requires careful interface management. Lifecycle planning should therefore include version control, traceability, and rollback procedures. At the research ecosystem level, the sustainability of photoplethysmographic signal processing depends on the availability of long-term benchmark datasets and openly documented evaluation protocols. A fragmented field in which each group uses private datasets and noncomparable metrics cannot accumulate reliable evidence about which architectural choices are robust. Shared benchmark initiatives, even when they involve deidentified data, can accelerate progress and reduce redundant data collection. Such initiatives must be designed with privacy protections and data governance standards that respect participant autonomy. In this sense, sustainability is not only an environmental or economic concern but also a property of the research infrastructure itself.

9. Future Directions

Future research directions should address several structural gaps in current deep learning-based photoplethysmographic artifact removal. First, multimodal sensor fusion beyond accelerometry warrants systematic exploration. Combining photoplethysmography with electrocardiography, impedance pneumography, or temperature signals may improve artifact separation, but it also raises synchronization and privacy challenges. The trade-off between additional information and increased system complexity must be evaluated through multidisciplinary studies that include clinical endpoints, not only waveform reconstruction error. Second, self-supervised and semi-supervised learning methods should be developed to reduce dependence on fully paired clean and corrupted signals. Real-world ambulatory data are abundant but often lack reliable ground truth. Self-supervised pretext tasks, such as reconstructing masked portions of the waveform or predicting motion-induced distortions, can learn useful representations without requiring extensive clinical annotation. However, these methods must be validated against downstream physiological accuracy to ensure that the learned representations preserve clinically meaningful information. Third, uncertainty estimation should become a standard component of artifact removal systems. A model that communicates when it is uncertain can prevent erroneous downstream decisions and support clinician trust. Uncertainty-aware architectures, including Monte Carlo dropout and deep ensembles, have shown promise in other biomedical domains and should be adapted to photoplethysmography.

Fourth, federated learning offers a pathway to improve model robustness across diverse populations without centralizing sensitive physiological data. Federated approaches can train a shared model across many institutions or devices while keeping raw waveforms local. This can increase demographic diversity and reduce privacy risk, but it introduces challenges related to data heterogeneity, communication efficiency, and model auditing. Research should examine how to combine federated training with fairness constraints and subgroup

performance monitoring. Fifth, the field needs standardized benchmarks that include multiple motion types, skin pigmentation levels, sensor placements, and photoplethysmographic sensor hardware. Such benchmarks should be developed through international consensus and should include transparent evaluation protocols. They would enable meaningful comparison across models and build the cumulative evidence needed for regulatory acceptance. Finally, future work should integrate artifact removal with clinical decision support workflows. Rather than treating denoising as a standalone preprocessing step, researchers should study how artifact removal errors propagate into diagnostic decisions and how to design safeguards that minimize patient harm. This requires collaboration among signal processing engineers, clinicians, regulatory scientists, and health system administrators. The goal is not to replace clinical judgment but to provide a reliable physiological signal foundation that supports equitable and accountable digital health.

10. Conclusion

Deep learning-based motion artifact removal for wearable photoplethysmographic signals offers considerable promise, but its value depends on more than raw denoising accuracy. This paper has argued that artifact removal should be understood as a sociotechnical infrastructure that spans optical sensing, model architecture, data governance, edge deployment, clinical validation, fairness, and policy. Architectural choices interact with energy and latency constraints, while data provenance and population representativeness shape algorithmic equity. Deployment models influence privacy, regulatory burden, and sustainability. A narrow focus on waveform reconstruction error obscures these systemic trade-offs and may produce systems that fail in vulnerable populations or under real-world conditions. A more robust approach integrates uncertainty estimation, continuous monitoring, and transparent governance, ensuring that learned denoisers are reliable across diverse users and settings. Future advances in multimodal fusion, self-supervised learning, federated training, and standardized benchmarking can strengthen the evidence base and support clinical translation. Ultimately, the goal is not merely to remove motion artifacts but to build trustworthy, sustainable, and equitable wearable cardiovascular monitoring infrastructure that serves a broad population.

References

1. Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), R1–R39.
2. Piwek, L., Ellis, D. A., Andrews, S., & Joinson, A. (2016). The rise of consumer health wearables: Promises and barriers. *PLoS Medicine*, 13(2), e1001953.
3. Ram, M. R., Madhav, K. V., Krishna, E. H., Komalla, N. R., & Reddy, K. A. (2012). A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter. *IEEE Transactions on Instrumentation and Measurement*, 61(5), 1445–1457.
4. Raghuram, M., Madhav, K. V., Krishna, E. H., & Reddy, K. A. (2014). Evaluation of wavelets for reduction of motion artifacts in photoplethysmographic signals. In 2014 10th International Conference on Information Sciences, Signal Processing and their Applications (ISSPA) (pp. 460–463). IEEE.
5. Reiss, A., Indlekofer, I., Schmidt, P., & Van Laerhoven, K. (2019). Deep PPG: Large-scale heart rate estimation with convolutional neural networks. *Sensors*, 19(14), 3079.

6. Castaneda, D., Esparza, A., Ghamari, M., Soltanpur, C., & Nazeran, H. (2018). A review on wearable photoplethysmography sensors and their potential future applications in health care. *International Journal of Biosensors & Bioelectronics*, 4(4), 195–202.
7. Temko, A. (2017). Accurate heart rate monitoring during physical exercises using PPG. *IEEE Transactions on Biomedical Engineering*, 64(9), 2018–2025.
8. Fukushima, H., Kawanaka, H., Bhuiyan, M. S., & Oguri, K. (2012). Estimating heart rate using wrist-type photoplethysmography and acceleration sensor while running. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 2901–2904). IEEE.
9. Sun, B., & Zhang, Z. (2015). Photoplethysmography-based heart rate monitoring using asymmetric least squares spectrum subtraction and Bayesian decision theory. *IEEE Sensors Journal*, 15(12), 7161–7168.
10. 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.
11. Lee, J., Matsumura, K., Yamakoshi, K., Rolfe, P., Tanaka, S., & Yamakoshi, T. (2013). Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 1724–1727). IEEE.
12. Bonomi, A. G., Schipper, F., Eerikäinen, L. M., Margarito, J., van Dinther, R., Muesch, G., de Morree, H. M., & Aarts, R. M. (2018). Atrial fibrillation detection using a novel cardiac ambulatory monitor based on photo-plethysmography at the wrist. *Journal of the American Heart Association*, 7(15), e009351.
13. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
14. Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., & Ranganath, R. (2020). A review of challenges and opportunities in machine learning for health care. *Clinical Pharmacology & Therapeutics*, 107(4), 803–812.
15. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453.
16. Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). *AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations*. *Minds and Machines*, 28(4), 689–707.
17. 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.
18. He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, 25(1), 30–36.
19. World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization.

20. Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E., & Valley, T. S. (2020). Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, 383(25), 2477–2478.
21. Fine, J. S., Branan, K. L., Rodriguez, A. J., Boonya-Ananta, T., Ajmal, A., Ramella-Roman, J. C., & McShane, M. J. (2021). Sources of inaccuracy in photoplethysmography for continuous cardiovascular monitoring. *Biosensors*, 11(4), 126.
22. Perez, M. V., Mahaffey, K. W., Hedlin, H., Rumsfeld, J. S., Garcia, A., Ferris, T., Balasubramanian, V., Russo, A. M., Rajmane, A., Cheung, L., Hung, G., Lee, J., Kowey, P., Talati, N., Nag, D., Gummidipundi, S. E., Beatty, A., Hills, M. T., Desai, S., Granger, C. B., Desai, M., & Turakhia, M. P. (2019). Large-scale assessment of a smartwatch to identify atrial fibrillation. *New England Journal of Medicine*, 381(20), 1909–1917.