

Multi-Task Deep Learning for Joint Heart Rate and Respiratory Rate Estimation from Photoplethysmography

Deepak Remen

Department of Computer Science, University of Central Florida, Orlando, FL, USA.
draman@ucf.edu

Aahesh Shetty

School of Computing, Clemson University, Clemson, SC, USA.
ashish.shetty549@clemson.edu

Mason M. Remos

Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
hellomason@buffalo.edu

Abstract

Photoplethysmography has become a central modality for noninvasive physiological monitoring because it is inexpensive, easily integrated into wearable devices, and suitable for continuous data acquisition. Estimating heart rate and respiratory rate from a single photoplethysmography stream is attractive for clinical and consumer health applications but remains difficult under motion artifacts, variable tissue perfusion, skin pigmentation differences, sensor placement changes, and pathophysiological diversity. This paper presents a system-level analysis of multi-task deep learning architectures for joint estimation of heart rate and respiratory rate from photoplethysmography. Rather than focusing narrowly on a single algorithm, the discussion examines structural trade-offs among shared feature extractors, task-specific prediction heads, temporal modeling strategies, and signal preprocessing stages. The paper reviews the evolution from classical signal processing and feature engineering to end-to-end learning, emphasizing how shared representations can improve sample efficiency and reduce overfitting when task labels are noisy or incomplete. It further analyzes infrastructure requirements for training and deployment, including data provenance, annotation quality, model monitoring, interpretability, bias mitigation, and regulatory alignment. Key architectural tensions include latency versus accuracy, centralized versus on-device inference, personalization versus generalization, and fairness across heterogeneous populations. The deployment discussion addresses energy consumption, sensor fusion, edge computing, and continuous model updating. Policy implications are considered in relation to clinical validation, patient safety, privacy, and algorithmic accountability. The paper concludes that multi-task deep learning offers a promising unified framework for cardiorespiratory monitoring, but its translational success depends on careful system design, rigorous governance, and robust evaluation beyond aggregate performance metrics.

Keywords

photoplethysmography; heart rate; respiratory rate; multi-task deep learning; wearable sensing; physiological monitoring; signal processing; system design.

1. Introduction

Photoplethysmography has become a foundational sensing modality in both clinical and consumer health monitoring because it enables continuous, low-cost, noninvasive observation of pulsatile blood volume changes [1]. A photoplethysmography waveform encodes not only the cardiac cycle but also respiratory modulation, vasomotor activity, and local tissue perfusion, making it an attractive single-source signal for multiple physiological estimates. However, the richness of the signal is accompanied by considerable fragility. Sensor placement, ambient light, skin pigmentation, tissue compression, and subject motion can distort the waveform in ways that are difficult to separate from true physiological variability [2]. These distortions are especially severe during physical activity, when motion-induced intensity changes overlap with the cardiac-synchronous component and undermine conventional peak detection or spectral methods [3].

The simultaneous estimation of heart rate and respiratory rate from a single photoplethysmography stream is a particularly demanding multi-task estimation problem because the two parameters occupy different time scales and frequency regions, and they interact with the underlying signal through distinct physiological mechanisms. Heart rate estimation typically depends on the accurate identification of pulse intervals or the dominant cardiac frequency, whereas respiratory rate estimation often relies on subtle amplitude, frequency, and baseline modulations of the photoplethysmography envelope. A joint model can exploit shared waveform structure while preserving task-specific inference pathways. This paper provides a system-level analysis of multi-task deep learning architectures for this problem, with emphasis on structural trade-offs, data governance, deployment constraints, robustness, fairness, and policy implications. It argues that architectural choices cannot be separated from the broader sociotechnical infrastructure in which wearable health monitoring systems operate.

2. Background and Related Work

The extraction of heart rate from photoplethysmography under motion has been extensively studied because ambulatory monitoring frequently occurs outside controlled laboratory environments. Early algorithms combined accelerometer-based noise references with spectral subtraction or adaptive filtering. Later methods formulated motion artifact rejection as a signal decomposition problem in which the underlying cardiac component is separated from intermittent noise. The importance of robust motion reference signals and temporal consistency constraints has been demonstrated in several studies that achieved accurate heart rate monitoring during intensive physical exercise [4]. More recently, learning-based models have further improved robustness by replacing explicit handcrafted filtering stages with learned feature extractors. Lightweight convolutional neural networks have been proposed for wrist-worn photoplethysmography heart rate estimation with constrained computational budgets, making them suitable for wearable processors [5]. Large-scale convolutional neural networks trained on extensive photoplethysmography datasets have also shown improved generalization across sensor and subject conditions, although they may require substantial training resources and careful regularization [6].

Respiratory rate estimation from photoplethysmography has a different analytical history. The respiratory influence on the photoplethysmogram arises through several interacting mechanisms, including respiratory sinus arrhythmia, respiratory-induced intensity variation, and respiratory-induced amplitude modulation. A comprehensive review of breathing rate estimation from electrocardiogram and photoplethysmogram identified these mechanisms as partially redundant, which supports the use of fused signal representations rather than relying

on a single modulation source [7]. Subsequent work demonstrated that respiratory rate can be estimated from pulse oximeter waveforms by combining multiple modulation features and using quality-aware fusion, particularly when individual features are intermittently unreliable [8]. These findings suggest that a multi-task system should preserve low-frequency respiratory modulation rather than treating it as baseline noise during preprocessing.

Deep neural architectures have also demonstrated strong performance in related physiological signal tasks. Cardiologist-level arrhythmia detection from ambulatory electrocardiograms illustrates the capacity of end-to-end learning to capture subtle temporal patterns in bio-signals [9]. In parallel, traditional feature-based models such as gradient boosted trees remain useful as interpretable baselines and for resource-constrained deployment scenarios in which handcrafted features are stable and well understood [10]. The contrast between learned and handcrafted representations is central to the design of multi-task photoplethysmography systems because the optimal preprocessing strategy for heart rate may conflict with the requirements of respiratory rate estimation.

3. System Architecture and Signal Processing Pipeline

A complete photoplethysmography-based cardiorespiratory monitoring system includes a sensing front-end, analog conditioning, analog-to-digital conversion, signal segmentation, preprocessing, model inference, postprocessing, and application-level fusion. Each stage introduces design choices that affect downstream model performance. The sensing front-end may include light sources at different wavelengths, photodetectors, and accelerometers. Multi-wavelength sensing can provide additional information about tissue oxygenation and perfusion, but it also increases energy consumption and hardware complexity. Segmentation parameters such as window length and overlap determine the amount of temporal context available to the model. Short windows reduce latency and computational cost but may not capture slow respiratory modulations, while long windows improve respiratory rate estimation at the expense of delayed outputs and higher memory usage.

Preprocessing is a particularly important structural decision. Conventional heart rate pipelines often apply bandpass filtering to suppress baseline wander and high-frequency noise. However, respiratory information frequently resides in lower-frequency modulations that such filters may remove. Adaptive notch filtration methods have been developed to recover photoplethysmographic signals during physical activity by dynamically rejecting motion-related frequency components while preserving adjacent physiological content [11]. This creates an architectural tension. A system that aggressively cleans the signal for heart rate estimation may destroy the respiratory modulation needed by the joint model. Conversely, a system that preserves the raw signal too fully may expose the network to motion artifacts that could be handled more efficiently by a dedicated filter. The optimal balance depends on the capacity of the learned feature extractor, the quality of motion references, and the target deployment environment.

From a system perspective, preprocessing can be viewed as an architectural resource that can be placed before or inside the model. Classical preprocessing modules are interpretable and can be validated independently, but they may introduce irreversible information loss. Learned preprocessing layers can adapt to the data distribution and preserve task-relevant features, but they are less transparent and may overfit to sensor-specific artifacts. Hybrid designs that combine light analog filtering, adaptive digital filtering, and learned waveform normalization often provide the best trade-off in practice. The system must also manage synchronization between photoplethysmography waveforms and reference signals such as electrocardiogram

or capnography during training, because misalignment can corrupt the learning signal more severely than sensor noise.

4. Multi-Task Deep Learning Model Design

Multi-task learning has a long history as a mechanism for improving generalization by sharing representations across related tasks [12]. The foundational argument is that training a single network on multiple related tasks can bias the learner toward representations that are more robust than those learned by independent task-specific models [13]. In the context of photoplethysmography, this is especially relevant because heart rate and respiratory rate are physiologically coupled but not deterministically related. Shared convolutional encoders can learn a common time-frequency representation, while separate task heads decode task-specific outputs. This structure reduces the number of parameters relative to deploying two separate models and encourages the shared representation to retain multiple aspects of the waveform.

The choice of temporal modeling layer also affects the trade-off between accuracy and computational cost. Recurrent networks can model sequential dependencies but may be difficult to parallelize on edge devices. Attention mechanisms provide an alternative by allowing the model to weight temporal contexts dynamically, and transformer-style attention has influenced many physiological time-series models [14]. Deep hierarchical models, in general, can learn progressively abstract features from raw or lightly processed waveforms, reducing the need for manual feature engineering [15]. However, deeper models require larger training datasets and more careful regularization to avoid overfitting to sensor-specific artifacts.

A central system challenge in multi-task learning is task conflict. Heart rate and respiratory rate may favor different temporal resolutions, and gradient updates from one task may interfere with the other. Architectural separation at later layers can reduce destructive interference while still allowing shared low-level features. Soft parameter sharing and task-specific gating mechanisms offer more flexibility but increase model size and complexity. The training objective must also balance task-specific losses without letting the higher-variance task dominate optimization. In practice, the best configuration depends on the amount of labeled data, the reliability of each reference signal, and the deployment constraints of the target device. A well-designed multi-task model can improve both sample efficiency and robustness, but a poorly balanced design may perform worse than two carefully tuned single-task models.

5. Training Infrastructure and Data Governance

Training large-scale multi-task photoplethysmography models requires curated datasets that represent diverse age groups, skin pigmentation, body mass index, perfusion states, medication effects, and motion patterns. Photoplethysmography datasets are often collected with reference devices such as chest-worn electrocardiogram and capnography or respiratory inductance plethysmography. Label noise and synchronization error are common because reference devices may have different time constants, attachment methods, and failure modes. The quality and provenance of these labels must be governed explicitly. Model cards provide structured documentation for reporting model performance across subgroups, intended use, and limitations, which is essential for clinical and consumer-facing systems [16]. Datasheets for datasets extend this governance to the data collection process, documenting inclusion criteria, collection conditions, known biases, and recommended use [17]. These practices are

critical because a model may exhibit excellent aggregate error rates while failing on underrepresented populations.

Scalable learning systems for health data can facilitate training across multiple institutions, but they must address heterogeneous data schemas, clinician annotation workflows, and privacy-preserving federated analysis [18]. Federated infrastructure allows model updates to be computed at local sites without direct sharing of raw health data, but it introduces additional challenges in versioning, heterogeneity, and security. Annotation workflows for respiratory rate are particularly difficult because reference respiratory signals can be noisy in free-living conditions. Data governance must therefore include explicit processes for label auditing, subject-level splitting, temporal splitting, and stress testing under distribution shift. Training infrastructure should support reproducible experiments, model lineage tracking, and audit trails for regulatory review. Without these mechanisms, the technical performance of a multi-task model cannot be translated into trustworthy deployments.

6. Deployment, Robustness and Fairness

Wearable deployment imposes severe computational, memory, energy, and thermal constraints. Multi-task models may share feature extraction to reduce redundant computation compared with two separate models, but task-specific heads and long temporal context may raise latency. Quantization, pruning, and neural architecture search can reduce model size, but these techniques may differentially degrade performance on low-prevalence subgroups or rare physiological patterns. Deployment engineers must therefore evaluate not only average error but also worst-case behavior under sensor displacement, low perfusion, motion bursts, and environmental interference. A guide to deep learning in healthcare emphasizes that clinical translation requires prospective validation, integration into clinician workflows, and attention to quality control rather than demonstration of retrospective accuracy alone [19].

Fairness in photoplethysmography-based monitoring is not only a technical concern but also an ethical one. Differences in skin pigmentation, skin thickness, vascular structure, and sensor contact can alter signal intensity and noise characteristics. If training data are biased toward healthy, light-skinned, or well-perfused populations, the resulting model may produce systematically larger errors for other groups. Responsible machine learning for health care requires that fairness, transparency, and accountability be addressed before deployment, including subgroup performance reporting and external validation on diverse populations [20]. Continuous monitoring after deployment is also necessary because sensor hardware revisions, user behavior, and seasonal changes can shift the data distribution. Drift detection, calibration checks, and postmarket surveillance should be integrated into the system architecture rather than treated as retrospective safety activities.

A further deployment consideration is personalization. A general model trained on population data may not capture individual physiological idiosyncrasies, such as unusual pulse morphology or respiratory patterns. On-device adaptation can improve personalization, but it raises questions about model versioning, error accumulation, and the loss of centralized oversight. System designers must decide whether personalization updates are stored locally, aggregated centrally, or restricted to a thin calibration layer. The chosen strategy has direct implications for privacy, equity, and regulatory compliance. A federated personalization approach can reduce raw data transfer while still enabling population-level learning, but it requires robust governance to prevent unstable or biased local updates.

7. Infrastructure Sustainability and Policy Implications

Sustainability considerations are increasingly important in large-scale health monitoring systems. Training large deep learning models consumes substantial energy, and continuous on-device inference drains wearable batteries. Multi-task architectures can reduce total energy by sharing parameters compared with separate heart rate and respiratory rate models, but larger shared networks may still be unsuitable for battery-powered devices. Model compression, efficient attention approximations, and adaptive sampling strategies can reduce energy demand, but each optimization must be evaluated for its impact on clinical utility and subgroup fairness. The lifecycle of wearable devices also introduces e-waste and data retention concerns, especially if devices are replaced frequently or if models require periodic retraining on resource-intensive cloud infrastructure.

Policy implications extend beyond technical performance. Regulatory frameworks for software as a medical device increasingly expect evidence of subgroup performance, postmarket surveillance, and controlled software versioning. A multi-task model that simultaneously estimates two vital signs may be subject to different risk classifications depending on clinical intended use and user population. Model updates that alter behavior require staged validation, and continuous learning systems must demonstrate that automated updates do not introduce harm. Governance structures should include data stewards, clinicians, engineers, and patient representatives to ensure that system-level decisions reflect clinical priorities and social values. Privacy is another central policy concern because photoplethysmography signals can reveal information about cardiovascular status, stress, and potentially disease states. Data minimization, consent management, and secure storage must be designed into the system rather than added after deployment.

The broader adoption of multi-task deep learning for cardiorespiratory monitoring also depends on the development of shared evaluation benchmarks that report not only mean absolute error but also reliability, calibration, latency, energy use, and subgroup performance. Such benchmarks would enable regulators, clinicians, and researchers to compare systems under realistic conditions and to identify failure modes that aggregate metrics conceal. The academic community and industry must collaborate with standards bodies to define acceptable performance envelopes for different use contexts, from wellness tracking to clinical triage. This sociotechnical alignment is as important as architectural innovation for the long-term success of photoplethysmography-based monitoring.

8. Conclusion

Multi-task deep learning offers a compelling framework for joint heart rate and respiratory rate estimation from photoplethysmography because it can exploit shared waveform structure while preserving task-specific inference pathways. This paper has examined the problem from a system-level perspective, emphasizing that model architecture cannot be isolated from preprocessing decisions, data governance, deployment infrastructure, robustness, fairness, and policy. The structural tension between aggressive motion artifact removal and the preservation of respiratory modulation illustrates the need for careful integration of signal processing and learned feature extraction. Shared representations can improve sample efficiency and reduce computational overhead, but task conflicts and imbalance must be managed through thoughtful architectural design. Training and deployment require robust data provenance, subgroup-level evaluation, and continuous monitoring to ensure that the system performs safely across heterogeneous populations. Energy consumption, device lifecycle sustainability, and regulatory compliance further shape the design space. Ultimately, the successful translation of joint cardiorespiratory estimation from photoplethysmography

will depend not only on algorithmic accuracy but also on the broader sociotechnical infrastructure that governs how such systems are built, evaluated, and maintained.

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. Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. *Current Cardiology Reviews*, 8(1), 14–25. <https://doi.org/10.2174/157340312801215782>
3. 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>
4. Temko, A. (2017). Accurate heart rate monitoring during physical exercises using PPG. *IEEE Transactions on Biomedical Engineering*, 64(9), 2017–2024. <https://doi.org/10.1109/TBME.2017.2676243>
5. Biswas, D., Simões-Capela, N., Van Hoof, C., & Van Helleputte, N. (2019). Heart rate estimation from wrist-worn PPG using a lightweight convolutional neural network. *Sensors*, 19(18), 3870. <https://doi.org/10.3390/s19183870>
6. 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. <https://doi.org/10.3390/s19143079>
7. Charlton, P. H., Birrenkott, D. A., Bonnici, T., Pimentel, M. A. F., Johnson, A. E. W., Alastruey, J., Tarassenko, L., Watkinson, P. J., Beale, R., & Clifton, D. A. (2018). Breathing rate estimation from the electrocardiogram and photoplethysmogram: A review. *IEEE Reviews in Biomedical Engineering*, 11, 2–20. <https://doi.org/10.1109/RBME.2017.2763681>
8. Karlen, W., Raman, S., Ansermino, J. M., & Dumont, G. A. (2013). Multiparameter respiratory rate estimation from the photoplethysmogram. *IEEE Transactions on Biomedical Engineering*, 60(7), 1946–1953. <https://doi.org/10.1109/TBME.2013.2246160>
9. Hannun, A. Y., Rajpurkar, P., Haghighpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1), 65–69. <https://doi.org/10.1038/s41591-018-0268-3>
10. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
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. Caruana, R. (1997). Multitask learning. *Machine Learning*, 28(1), 41–75.
<https://doi.org/10.1023/A:1007379606734>
13. Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.
14. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* 30 (pp. 5998–6008).
15. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
<https://doi.org/10.1038/nature14539>
16. 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. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229).
<https://doi.org/10.1145/3287560.3287596>
17. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, 64(12), 86–92. <https://doi.org/10.1145/3458723>
18. 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>
19. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
20. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine*, 25(9), 1337–1340. <https://doi.org/10.1038/s41591-019-0548-6>