Uncertainty-Aware Vital Sign Estimation for Reliable Wearable Healthcare Applications

Adrian Andrews
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
adrian118@missouri.edu

Jean R. Zimmerman
Department of Computer Science, George Mason University, Fairfax, VA, USA.
jzimmerman@gmu.edu

Andres Makinen
School of Computing, Clemson University, Clemson, SC, USA.
andresm@clemson.edu

Abstract

Wearable healthcare systems increasingly produce continuous streams of physiological data for clinical screening, remote patient monitoring, and personal health management. However, the translation of raw sensor signals into reliable vital sign estimates remains difficult outside controlled laboratory settings because motion artifacts, sensor displacement, environmental variability, and anatomical differences introduce substantial uncertainty into the inference pipeline. Deterministic models often generate point estimates without communicating the confidence that should accompany each prediction, which can lead to overinterpretation, false alarms, or missed clinical deterioration. This paper examines uncertainty-aware vital sign estimation from a systems perspective, emphasizing architectural, infrastructural, governance, deployment, and policy considerations rather than algorithmic detail alone. It discusses how uncertainty quantification can be integrated across wearable sensing layers, edge computing nodes, clinical data platforms, and organizational workflows. The paper further explores structural trade-offs around real-time processing, energy consumption, data quality assurance, interoperability, and privacy preservation. Robustness and fairness are addressed as system properties that must be actively engineered across heterogeneous populations and use contexts. Sustainability challenges associated with device lifecycles, model drift, and continuous learning are considered alongside regulatory and ethical implications. By situating uncertainty-aware inference within the broader socio-technical infrastructure of wearable healthcare, the paper argues that reliable vital sign estimation requires coordinated design decisions spanning sensors, algorithms, data governance, and clinical integration. Future research should prioritize standardized uncertainty reporting, calibration under realistic deployment conditions, and governance mechanisms that make confidence information interpretable and actionable for patients, clinicians, and algorithmic oversight systems.

Keywords

wearable healthcare, uncertainty quantification, vital sign estimation, physiological monitoring, trustworthy artificial intelligence, health data governance.

1. Introduction

Wearable devices now provide continuous access to physiological parameters such as heart rate, respiration rate, oxygen saturation, and activity-related indicators, enabling a shift from episodic clinical measurement toward longitudinal health monitoring in everyday environments [1]. This shift creates significant opportunities for early detection of deterioration, personalized care, and large-scale public health surveillance. At the same time, wearable measurements are inherently more variable than controlled bedside measurements because users move through heterogeneous physical and social environments, sensors are subject to placement variation, and device hardware differs widely across manufacturers and product generations. These conditions make uncertainty a central rather than peripheral concern for reliable wearable healthcare applications. Without explicit uncertainty awareness, a vital sign estimate can present a misleading impression of accuracy even when the underlying signal is corrupted, ambiguous, or only partially informative.

Uncertainty-aware inference offers a conceptual and architectural response to this problem by quantifying the degree of confidence associated with each estimate, distinguishing between noise inherent in the measurement process and limitations in the model itself [2]. In wearable photoplethysmography, for example, motion artifacts during physical activity can produce rapid changes in signal morphology that resemble physiological variability, and algorithms must decide whether to report an estimate or suppress it. Robust extraction under varying activity intensity has been a longstanding challenge because heart rate, respiration, and oxygen saturation interact with movement in complex ways [3]. Rather than treating uncertainty as an afterthought, reliable systems should place it at the center of the inference architecture, allowing downstream clinical workflows to interpret estimates in relation to their confidence, provenance, and context.

This paper analyzes uncertainty-aware vital sign estimation as a systems problem involving sensing infrastructure, computational architecture, data governance, deployment strategy, fairness, sustainability, and regulatory oversight. The objective is not to provide a technical derivation of a specific algorithm but to examine how uncertainty is produced, communicated, and governed across the entire wearable healthcare stack. The analysis draws on cross-domain perspectives from biomedical signal processing, edge computing, health informatics, responsible artificial intelligence, and organizational risk management.

2. System-Level Challenges in Wearable Vital Sign Estimation

Wearable vital sign estimation depends on the interpretation of noisy physiological signals that are affected by optical coupling, tissue composition, blood perfusion, and environmental illumination, particularly in photoplethysmographic sensing [4]. The same underlying physiological state can produce different signal appearances across skin types, sensor placements, and physical activity levels. These variations create uncertainty that cannot be fully eliminated by improved hardware alone because some variability is inherent to the measurement context. A system-level perspective therefore treats uncertainty as a property that propagates from the sensor interface through preprocessing, feature extraction, model inference, and clinical interpretation.

Modern uncertainty quantification approaches offer practical mechanisms for representing this variability. Bayesian approximations such as Monte Carlo dropout provide a way to estimate model uncertainty without requiring a full Bayesian treatment of every neural network component [5]. Deep ensembles, which combine multiple independently trained models, have been shown to produce well-calibrated predictive distributions in many application domains [6]. Conformal prediction offers an alternative framework that provides

distribution-free guarantees about the coverage of prediction intervals, which is particularly appealing for safety-critical physiological monitoring because it does not rely on strong parametric assumptions [7]. These methods can be embedded within wearable inference pipelines to generate confidence intervals or to flag estimates that fall below a reliability threshold.

However, uncertainty estimation is itself vulnerable to device and context heterogeneity. Wearable systems vary in sampling rates, sensor quality, firmware behavior, and motion artifact characteristics, which means that a model trained on one device population may not produce calibrated uncertainties on another [8]. Recent work has demonstrated that vital signs can be extracted rapidly during varying physical activity intensity, but such capability must be paired with continuous signal quality assessment and confidence quantification to avoid generating numerically plausible yet clinically unreliable outputs [9]. This integration is not simply a modeling task; it requires architectural decisions about where uncertainty is computed, how it is represented, and how it influences downstream action.

3. Architectural Strategies for Uncertainty-Aware Inference

A layered architecture is often appropriate for wearable healthcare systems because it allows different computational capabilities and latency constraints to be balanced across sensing nodes, edge devices, and cloud services. Edge computing reduces the need to transmit raw physiological data continuously, thereby improving privacy, lowering bandwidth demand, and enabling real-time feedback [10]. At the same time, edge devices have limited energy and computational capacity, which restricts the complexity of uncertainty-aware models. Cloud-based processing can support larger ensembles and more sophisticated calibration procedures, but it introduces latency, connectivity dependence, and additional privacy risks. A well-designed system therefore distributes uncertainty computation according to clinical urgency, data sensitivity, and available computational resources.

Sensor fusion plays a critical role in improving the robustness of vital sign estimation under uncertain conditions. Combining photoplethysmography with accelerometry, electrocardiography, or temperature sensing can provide complementary evidence that reduces ambiguity when one sensor is disrupted by motion or environmental interference [11]. Uncertainty-aware fusion architectures can use signal quality indices to gate the contribution of each sensor, ensuring that corrupted channels do not dominate the final estimate. Signal quality assessment is essential because a vital sign algorithm cannot reliably interpret data from a misplaced or degraded sensor without first understanding whether the signal itself is trustworthy [12]. By propagating quality indicators alongside physiological estimates, systems can produce contextualized confidence values that support clinical reasoning.

The decision to compute uncertainty locally or remotely has direct implications for reliability. Local computation can reduce the risk of network failures but may produce less sophisticated uncertainty estimates. Remote computation can improve calibration but introduces dependency on external infrastructure that may not be available in all care settings. Hybrid approaches can use lightweight local models for immediate screening and more complex cloud models for periodic recalibration. The architectural trade-offs must be evaluated not only in terms of accuracy but also in terms of resilience, energy consumption, and alignment with clinical workflow requirements.

4. Data Infrastructure and Signal Quality Governance

Reliable wearable healthcare depends on data infrastructure that captures not only physiological measurements but also metadata about device configuration, sampling conditions, firmware version, signal quality, and user activity. Such metadata are essential for interpreting uncertainty because they allow the system to distinguish between estimates derived from clean signals and those derived from ambiguous or partial data. Interoperability standards, including health data exchange frameworks, provide a foundation for representing this contextual information consistently across institutions and software platforms [13]. Without standardized metadata, uncertainty estimates may be lost or misinterpreted when data move from wearable devices to clinical dashboards or electronic health records.

Privacy regulation further shapes the data infrastructure for uncertainty-aware systems. Wearable health data are sensitive because they can reveal information about behavior, movement, sleep, and physiological status over time. The General Data Protection Regulation establishes obligations around lawfulness, transparency, data minimization, and purpose limitation that affect how uncertainty-related metadata can be collected and processed [14]. System designers must ensure that the additional data needed for uncertainty quantification, such as raw sensor diagnostics or motion context, are governed under the same privacy protections as the vital signs themselves. Privacy-preserving processing at the edge may reduce the need to transmit identifiable raw data while still allowing meaningful uncertainty estimates to be computed locally.

Data drift and dataset shift present ongoing challenges for uncertainty calibration across real-world deployments. A model trained on supervised data collected from healthy volunteers in controlled conditions may not maintain calibrated uncertainty when applied to older adults, patients with compromised circulation, or users in hot and humid environments [15]. Governance mechanisms should therefore include continuous monitoring of signal distributions, periodic recalibration, and documentation of the populations and contexts represented in training data. Without such infrastructure, uncertainty estimates can become overconfident precisely when they are most needed.

5. Deployment Considerations in Heterogeneous Healthcare Environments

Deploying wearable uncertainty-aware systems in real healthcare environments requires attention to clinical workflow, device usability, and institutional variation. Wearable sensors have been widely used in athletic and wellness contexts, but clinical deployment involves different expectations around accuracy, alarm thresholds, and liability [16]. A system that reports uncertainty intervals may be more informative than a deterministic monitor, but only if clinicians understand how to interpret and act on that information. The introduction of uncertainty awareness therefore requires training, interface design, and decision support that tailor confidence information to clinical tasks.

Regulatory expectations also shape deployment decisions. Software that provides diagnostic or therapeutic recommendations is increasingly subject to regulatory oversight, particularly when machine learning methods are involved. The United States Food and Drug Administration has articulated an action plan for artificial intelligence and machine learning-based software as a medical device, emphasizing transparency, real-world performance monitoring, and lifecycle management [17]. Uncertainty-aware systems must be designed with these regulatory expectations in mind, including the ability to document model behavior, monitor calibration drift, and justify risk controls when uncertainty exceeds acceptable thresholds. Deployment should not be treated as a one-time event but as a continuous process of monitoring, recalibration, and safety assessment.

Heterogeneity across healthcare organizations further complicates deployment. A large academic medical center may have the technical staff and data infrastructure to interpret model confidence scores, while a small outpatient clinic may depend on simplified alerts and clear escalation paths. Uncertainty-aware systems must therefore support multiple presentation modes, from detailed clinician dashboards to patient-facing messages that communicate when a measurement may be unreliable. The structural requirement is not merely algorithmic accuracy but organizational fit, which includes clear roles, accountable oversight, and integration with existing patient safety processes.

6. Robustness, Fairness, and Clinical Risk Management

Robustness in wearable vital sign estimation means maintaining acceptable performance under diverse operational conditions, including sensor variation, motion intensity, environmental interference, and demographic differences. Fairness is closely related because performance and uncertainty may vary systematically across groups defined by skin tone, age, body composition, or underlying health conditions. If uncertainty estimates are not calibrated across these groups, the system may generate more frequent false alarms or missed events for some populations while appearing reliable for others [18]. Fairness therefore requires explicit evaluation of uncertainty calibration across stratified subgroups rather than relying on aggregate performance metrics alone.

Clinical risk management benefits from uncertainty information because it can enable graded responses. When confidence is high and the vital sign is normal, the system may remain silent. When confidence is low, the system can recommend a repeat measurement, sensor repositioning, or manual review. This approach can reduce alarm fatigue and improve trust in automated monitoring. However, trust is fragile and depends on whether clinicians perceive the uncertainty system as accurate, interpretable, and consistent with their own clinical judgment [19]. If uncertainty estimates are poorly calibrated or presented without context, clinicians may ignore them or develop an unjustified reliance on the technology.

Transparency mechanisms such as model cards and datasheets support robust risk management by documenting intended use, model limitations, evaluation results, and known failure modes [20]. In healthcare settings, such documentation should be integrated with organizational governance frameworks that define accountability for algorithmic decisions. Responsible machine learning in medicine requires not only technical validation but also institutional structures that can respond to errors, support clinicians in challenging cases, and protect patients from cascading failures [21]. Uncertainty-aware systems are more likely to be dependable if their confidence outputs are accompanied by clear documentation of how those outputs were produced and under what conditions they should be trusted.

7. Sustainability and Lifecycle Perspectives

Sustainability considerations are often neglected in discussions of machine learning for wearable healthcare, yet they are directly relevant to system design and long-term reliability. Continuous sensing and frequent model inference consume energy, which shortens battery life and affects user adherence. Energy-efficient context recognition strategies can reduce the computational burden of wearable monitoring while preserving the ability to identify critical physiological events [22]. Uncertainty-aware systems must balance the benefits of frequent confidence updates against the energy cost of running multiple models or ensembles on resource-constrained devices.

The environmental footprint of wearable devices also extends beyond operational energy use to manufacturing, replacement, and electronic waste. Lifecycle assessments of consumer electronics have shown that the environmental burden varies substantially depending on materials, production processes, and device longevity [23]. Frequent model updates that require newer hardware can shorten the useful life of devices, creating tension between algorithmic improvement and sustainability. System designers should consider modular hardware, over-the-air updates, and backward-compatible uncertainty modules that can be integrated without forcing premature device replacement.

From a data lifecycle perspective, models may degrade as patient populations, sensor technologies, and care protocols evolve. Continuous learning can improve performance, but it also introduces risks of catastrophic forgetting, concept drift, and unmonitored changes in behavior. Sustainable governance requires versioned models, archived datasets, and periodic audits that compare current system behavior with prior validation evidence. Such practices make it possible to maintain uncertainty calibration over time without assuming that a model validated at one moment will remain safe indefinitely.

8. Policy, Regulatory, and Organizational Implications

Policy development for wearable healthcare must address the growing role of probabilistic reasoning in clinical decision support. Ethical frameworks for artificial intelligence emphasize beneficence, non-maleficence, autonomy, and justice, but operationalizing these principles in wearable systems requires concrete guidance about uncertainty thresholds, alert escalation, and documentation [24]. Policymakers should consider whether clinical guidelines need to specify acceptable levels of uncertainty before an automated system recommends action, and how such thresholds should vary across diagnostic, monitoring, and wellness use cases.

Equity is a central policy concern because wearable healthcare is deployed within existing social and economic inequalities. Devices may be less accessible to low-income populations, less accurate across diverse skin tones, or less supported by digital literacy resources. Well-intentioned informatics interventions can inadvertently widen disparities if they assume uniform access, engagement, and technical infrastructure [25]. Policy interventions should promote inclusive design, representative validation data, and reimbursement mechanisms that do not penalize populations for whom uncertainty may be higher due to historical underrepresentation.

Organizational governance must link uncertainty-aware algorithms to accountable human decision-makers. Hospitals, insurers, and device manufacturers need clear protocols for investigating missed events, unexpected confidence values, and calibration failures. These protocols should be part of broader patient safety systems rather than isolated artificial intelligence oversight committees. By embedding uncertainty governance into existing quality improvement and risk management structures, organizations can create a sustainable path toward reliable wearable healthcare that respects both technical limitations and clinical responsibilities.

9. Future Directions and Research Agenda

Future research should move beyond isolated algorithm benchmarks toward integrated evaluations of uncertainty-aware wearable systems in realistic settings. Longitudinal studies are needed to examine how confidence estimates change over days, across seasonal conditions, and during clinical deterioration. Calibration should be assessed under dataset shift, sensor replacement, and firmware updates, with an emphasis on subgroup-specific

performance. Standardized reporting of uncertainty metrics would support comparison across devices and algorithms, but such standards must be developed with input from engineers, clinicians, regulators, and patient representatives.

Research is also needed on the human factors of uncertainty communication. Clinicians and patients may interpret probability statements differently depending on numeracy, task urgency, and interface design. Studies should examine how confidence intervals, categorical risk labels, and visual representations affect trust, decision speed, and clinical outcomes. The development of adaptive interfaces that present uncertainty only when necessary may reduce cognitive overload while preserving safety.

The governance of uncertainty-aware systems warrants comparative research across jurisdictions. Different regulatory regimes approach algorithmic transparency and medical device oversight in distinct ways, and these differences may shape how wearable manufacturers implement and document uncertainty quantification. Cross-national studies can identify practices that support reliability without imposing excessive administrative burden. Such evidence would help align technical innovation with public policy, enabling wearable healthcare to achieve its potential as a trustworthy component of modern health systems.

10. Conclusion

Uncertainty-aware vital sign estimation has the potential to make wearable healthcare applications more reliable, transparent, and clinically useful. However, achieving this potential requires a systems perspective that integrates signal processing advances with architectural design, data governance, deployment strategy, fairness assessment, sustainability planning, and regulatory compliance. Deterministic models that ignore uncertainty may perform well in controlled evaluations while failing in the messy conditions of everyday life. Systems that explicitly represent and communicate uncertainty can support more cautious interpretation, more effective triage, and more accountable clinical workflows. The structural challenge is not simply to add uncertainty outputs to existing algorithms but to redesign the entire wearable healthcare stack so that confidence information is treated as a first-class object throughout sensing, inference, communication, and oversight. As wearable devices become more embedded in care delivery, uncertainty-aware design should be viewed as a core requirement for trustworthy health technology rather than an optional enhancement.

References

1. Majumder, S., Mondal, T., & Deen, M. J. (2017). Wearable sensors for remote health monitoring. Sensors, 17(1), 130.

2. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? Advances in Neural Information Processing Systems, 30, 5574-5584.

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.

4. Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement, 28(3), R1-R39.

5. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. Proceedings of the 33rd International Conference on Machine Learning, 1050-1059.

6. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems, 30, 6402-6413.

7. Angelopoulos, A. N., & Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. arXiv preprint arXiv:2107.07511.

8. Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T. S., Kjaergaard, M. B., Dey, A., Sonne, T., & Jensen, M. M. (2015). Smart devices are different: Assessing and mitigating mobile sensing heterogeneity for activity recognition. Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, 127-140.

9. Zheng, X., Dwyer, V. M., Barrett, L. A., Derakhshani, M., & Hu, S. (2023). Rapid vital sign extraction for real-time opto-physiological monitoring at varying physical activity intensity levels. IEEE Journal of Biomedical and Health Informatics, 27(7), 3107-3118.

10. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637-646.

11. Nemati, S., Malhotra, A., & Clifford, G. D. (2010). Data fusion for improved respiration rate estimation. EURASIP Journal on Advances in Signal Processing, 2010, 926305.

12. Orphanidou, C., Bonnici, T., Charlton, P., Clifton, D., Vallance, D., & Tarassenko, L. (2015). Signal-quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications to wireless monitoring. IEEE Journal of Biomedical and Health Informatics, 19(3), 832-838.

13. Benson, T., & Grieve, G. (2021). Principles of health interoperability: FHIR, HL7 and SNOMED CT (4th ed.). Springer.

14. European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Union, L119, 1-88.

15. Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., & Lawrence, N. D. (2009). Dataset shift in machine learning. MIT Press.

16. Seshadri, D. R., Li, R. T., Voos, J. E., Rowbottom, J. R., Alfes, C. M., Zorman, C. A., & Drummond, C. K. (2019). Wearable sensors for monitoring the internal and external workload of the athlete. NPJ Digital Medicine, 2, 71.

17. U.S. Food and Drug Administration. (2021). Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan. U.S. Food and Drug Administration.

18. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866-872.

19. Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial intelligence and human trust in healthcare: Focus on clinicians. Journal of Medical Internet Research, 22(6), e15154.

20. 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.

21. 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.

22. Rault, T., Bouabdallah, A., Challal, Y., & Marin, F. (2017). A survey of energy-efficient context recognition systems using wearable sensors for healthcare applications. Pervasive and Mobile Computing, 37, 23-44.

23. Andrae, A. S. G., & Andersen, O. (2010). Life cycle assessments of consumer electronics: Are they consistent? International Journal of Life Cycle Assessment, 15(8), 827-836.

24. Floridi, L., Cowls, 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.

25. Veinot, T. C., Mitchell, H., & Ancker, J. S. (2018). Good intentions are not enough: How informatics interventions can worsen inequality. Journal of the American Medical Informatics Association, 25(8), 1080-1088.