

# **Anomaly Detection in Long-Term Wearable Cardiovascular Monitoring Using Uncertainty-Aware Deep Learning**

Jereme Boehan

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

Rohan D. Garg

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

## **Abstract**

Long-term wearable cardiovascular monitoring generates continuous streams of electrocardiographic and photoplethysmographic data that can support early detection of arrhythmias and other cardiac anomalies. Deep learning models have shown strong diagnostic performance in controlled settings, but deployment in ambulatory environments introduces significant uncertainty due to motion artifacts, sensor displacement, population heterogeneity, and distributional drift over time. This paper provides a system-level examination of anomaly detection in wearable cardiovascular monitoring using uncertainty-aware deep learning. It argues that predictive uncertainty should be treated as a systemic design requirement rather than a localized algorithmic property. The analysis covers architectural considerations, data governance, edge-cloud trade-offs, calibration, robustness, fairness, clinical accountability, sustainability, and policy implications. Uncertainty-aware mechanisms can reduce silent failures, support staged clinical escalation, and improve trust in longitudinal monitoring systems. However, these benefits depend on organizational integration, transparent data practices, and regulatory alignment. The paper synthesizes existing research on deep learning, uncertainty quantification, wearable health systems, and responsible clinical machine learning. It concludes that future systems must couple uncertainty estimation with operational workflows that can interpret and act on uncertain outputs. This requires continued collaboration among engineers, clinicians, regulators, and health system administrators to ensure that wearable cardiovascular monitoring becomes safe, equitable, and sustainable outside controlled research environments.

## **Keywords**

cardiovascular monitoring; wearable sensors; deep learning; uncertainty quantification; anomaly detection; clinical decision support; edge computing; data governance.

## **1. Introduction**

The expansion of wearable sensing technologies has transformed cardiovascular monitoring from an episodic clinical activity into a continuous ambulatory practice. Electrocardiographic and photoplethysmographic signals can now be collected over extended periods, producing high-volume data that exceed the capacity of conventional threshold-based review. Deep learning has become a central analytical approach for extracting complex physiological patterns from these large datasets [1]. In cardiology, deep neural networks have been shown

to detect arrhythmias in ambulatory electrocardiograms with performance comparable to experienced clinicians [2]. Related work has demonstrated that artificial intelligence can identify cardiac contractile dysfunction from routine electrocardiograms [3], and that deep networks can automate twelve-lead electrocardiogram interpretation across a range of conditions [4]. Convolutional architectures have also produced strong arrhythmia detection performance in single-lead recordings [5]. These developments suggest that continuous algorithmic surveillance could extend diagnostic attention beyond infrequent clinic visits and enable earlier intervention.

However, long-term wearable monitoring is fundamentally different from retrospective benchmark evaluation. The ambulatory environment introduces motion artifacts, changes in skin contact, sensor aging, and diverse patient behaviors that are much harder to control than hospital-based recordings. Physiological signals may be corrupted, incomplete, or produced under conditions that were not represented during model training. In such settings, deterministic deep learning models can be overconfident, and their errors may remain invisible until they cause clinical harm. Anomaly detection therefore requires not only accurate classification but also reliable estimates of uncertainty that can guide further investigation. The central argument of this paper is that uncertainty awareness must be understood as a property of the entire monitoring system, spanning hardware, data infrastructure, model design, clinical workflows, and governance mechanisms. The following sections examine these dimensions from a systems perspective, emphasizing structural trade-offs rather than isolated model performance.

## **2. Background and Related Work**

Wearable health monitoring has been the subject of extensive research for several decades. Early surveys emphasized the integration of physiological sensing, preprocessing, feature extraction, and decision support within wearable architectures [6]. More recent work has documented the shift toward deep learning in healthcare, where hierarchical representations learned from raw signals can replace manually engineered features [7]. Despite this progress, clinical implementation remains difficult because of data heterogeneity, label noise, interoperability constraints, and the need for continuous evaluation across different populations [8]. These challenges are especially acute in cardiovascular monitoring, where abnormal events are rare, signal quality varies widely, and the cost of missed detection is high.

Uncertainty quantification has emerged as an important complement to deep learning accuracy. Bayesian approximations such as dropout-based inference provide a practical mechanism for representing model uncertainty without abandoning standard neural network training [9]. The distinction between aleatoric uncertainty, which arises from noise in the data, and epistemic uncertainty, which arises from limited knowledge about the model, has been formalized and applied to computer vision [10]. The same conceptual distinction is relevant to physiological signals, where sensor noise and physiological variability represent aleatoric sources, while distributional shift and limited clinical labels represent epistemic sources. Deep ensembles provide another robust approach to predictive uncertainty by aggregating multiple independently trained models [11]. Calibration methods further correct models that produce systematically overconfident probabilities [12]. More recently, single deterministic networks have been designed to estimate uncertainty through distance-sensitive representation learning, offering a computationally efficient alternative to ensembles and Bayesian approximations [13]. These methods establish a technical foundation for uncertainty-aware monitoring but do not by themselves address the broader system requirements of wearable deployment.

### **3. System Architecture for Wearable Cardiovascular Monitoring**

A comprehensive wearable cardiovascular monitoring architecture must coordinate sensing, signal conditioning, local inference, communication, and clinical review. Photoplethysmographic signals are particularly vulnerable to motion corruption during physical activity, and adaptive filtering approaches are necessary to recover usable waveforms before downstream analysis [14]. Therefore, signal quality estimation should be treated as a first-class architectural function rather than as an optional preprocessing stage. An uncertainty-aware signal quality gate can be placed at the edge of the network to suppress corrupted inputs and forward only observations with sufficient confidence. This design reduces unnecessary transmission and protects central models from systematically degraded inputs.

The placement of computation across wearable nodes, gateways, and cloud services involves important trade-offs. On-device processing reduces latency and strengthens privacy, but it is constrained by memory, battery capacity, and the difficulty of frequent model updates. Cloud-based processing enables larger models and centralized auditing, but it depends on connectivity and creates additional risks in data aggregation. A hybrid architecture can balance these concerns by using lightweight uncertainty estimators on the device to triage events, while reserving detailed cloud analysis for cases with high epistemic uncertainty or disagreement among local models. This staged approach aligns with responsible machine learning frameworks that emphasize auditability, escalation protocols, and graduated autonomy [15]. The convergence of human and artificial intelligence in high-performance medicine similarly depends on reliable asynchronous monitoring rather than isolated algorithmic accuracy [16]. Furthermore, delivering clinical impact requires attention to workflow integration, interoperability, and continuous operational evaluation [17]. These system-level factors often determine whether an algorithm is adopted more than incremental improvements in sensitivity or specificity.

### **4. Uncertainty-Aware Deep Learning in Cardiac Anomaly Detection**

Uncertainty-aware anomaly detection requires that uncertainty estimates be actionable within clinical workflows. Aleatoric uncertainty can trigger signal reacquisition, filtering, or sensor adjustment, while epistemic uncertainty can prompt expert review or additional diagnostic testing. However, uncertainty is not a substitute for explainability. In medical artificial intelligence, explainability supports clinician understanding but cannot compensate for statistical unreliability [18]. Safe and reliable machine learning frameworks emphasize explicit modeling of failure modes, subgroup performance, and deployment conditions rather than relying on aggregate accuracy alone [19]. Uncertainty estimates offer one such reliability signal, but they must be integrated with evidence about context, patient history, and data provenance.

From an architectural perspective, deep residual networks have become a common backbone for extracting hierarchical features from physiological waveforms because they support deep representations while mitigating optimization difficulties [20]. When uncertainty mechanisms are added to these architectures, system designers face trade-offs among Bayesian approximations, deep ensembles, and deterministic uncertainty methods. Deep ensembles generally provide stable predictive distributions but multiply training and inference costs. Dropout-based Bayesian approximations introduce modest overhead but may require careful calibration and can be sensitive to implementation details. Deterministic uncertainty approaches can be more efficient, but their performance depends on representation

regularisation. The choice among these methods should be evaluated under realistic motion artifact conditions, with explicit attention to calibration drift, latency, and memory constraints. An uncertainty mechanism that performs well on curated datasets may become unreliable when sensors age, patient demographics shift, or recording contexts change.

## **5. Data Governance and Quality Infrastructure**

Long-term wearable monitoring depends on data governance structures that extend beyond technical preprocessing. Continuous physiological data are often unlabeled, context-dependent, and collected under heterogeneous consent frameworks. Data governance must address provenance, consent, de-identification, retention, and the conditions under which algorithms may repurpose clinical data. Without clear governance, models may be trained on biased or incomplete distributions and then deployed across populations for which they were not validated. This problem is particularly significant for cardiovascular monitoring because physiological signals vary with age, sex, body composition, skin tone, medication use, and comorbid conditions.

A robust data quality infrastructure should maintain longitudinal metadata about sensor type, firmware version, wear position, activity context, and signal quality indicators. Such metadata allow uncertainty estimates to be interpreted in relation to the conditions that produced the data. They also support post-market surveillance and model updating. When uncertainty signals indicate distributional drift, governance mechanisms should trigger review rather than silently continuing operation. This approach treats uncertainty as part of an institutional memory that can detect emerging risks. It also supports accountability by associating algorithmic outputs with the data conditions under which they were generated. Without such infrastructure, uncertainty-aware models may generate useful signals that clinical organizations cannot interpret or trust.

## **6. Deployment and Edge-Cloud Trade-offs**

Deployment decisions shape the reliability, privacy, and sustainability of wearable monitoring systems. Edge-first architectures keep sensitive physiological data closer to the patient and reduce dependence on network connectivity. They are especially valuable for continuous monitoring in rural or low-resource settings where connectivity may be intermittent. However, edge devices have limited computational capacity, and uncertainty estimation methods that require large ensembles or repeated stochastic forward passes can exceed these limits. Model compression, quantization, and architecture search can reduce computational load, but they may also degrade calibration if not explicitly evaluated.

Cloud-centric architectures enable central monitoring, frequent model updates, and detailed retrospective analysis. They also facilitate federated learning across institutional boundaries, although federated learning introduces additional governance and security challenges. A balanced deployment may use device-level models for immediate anomaly screening and cloud-based models for periodic or event-driven review. The choice of deployment topology should consider not only technical performance but also energy consumption, equipment lifespan, patient acceptability, and the resilience of the overall health system. Sustainability is a central but often underappreciated concern, since continuous monitoring can increase electronic waste and energy demand if devices and model updates are not managed carefully.

## **7. Fairness, Robustness, and Clinical Accountability**

Fairness in wearable cardiovascular monitoring requires attention to how model performance varies across demographic and clinical subgroups. If training data overrepresent certain populations, anomaly detectors may perform poorly for underrepresented groups. Uncertainty estimates can help identify these performance gaps, but they can also mask them if high uncertainty is systematically associated with specific populations and leads to unequal referral patterns. Fairness audits should therefore examine whether uncertainty-based escalation produces differential outcomes. This requires linking model outputs to clinical decisions and long-term health outcomes rather than evaluating algorithmic fairness in isolation.

Robustness is closely related to fairness. A robust system should maintain acceptable performance under sensor degradation, motion artifact, and changing patient behaviors. Uncertainty-aware models can improve robustness by flagging out-of-distribution inputs, but they are not inherently robust. Adversarial perturbations, calibration drift, and rare physiological events can still produce misleading outputs. Clinical accountability requires that organizations define clear lines of responsibility for algorithmic decisions. When an anomaly is missed or a false alarm leads to unnecessary treatment, accountability should rest within a defined governance structure that includes model developers, clinicians, and health system administrators. Uncertainty quantification supports accountability by providing evidence about the model's confidence at the time of a decision, but it cannot replace institutional responsibility.

## **8. Sustainability and Policy Implications**

The long-term sustainability of wearable cardiovascular monitoring depends on aligning technical innovation with policy frameworks that protect patients and promote equitable access. Regulatory pathways must adapt to continuous-learning systems that may update over time. Traditional approvals based on fixed model snapshots are poorly suited to algorithms that evolve in response to new data. Policy mechanisms should require evidence of calibration, uncertainty behavior, and subgroup performance across realistic deployment conditions. Post-market surveillance should monitor not only adverse events but also data quality, model drift, and disparities in access or outcomes.

From a policy perspective, uncertainty-aware systems can support risk-based regulation by allowing lower-risk triage functions to operate with greater autonomy while requiring human review for high-uncertainty or high-acuity cases. This graduated approach may reduce clinician burden without compromising safety. However, it also raises questions about liability, informed consent, and the digital divide. Patients with limited digital literacy or connectivity may be excluded from the benefits of continuous monitoring. Policymakers should therefore consider infrastructure investments, interoperability standards, and reimbursement models that support equitable deployment. Sustainability requires balancing the energy costs of cloud computing and frequent model updates against the potential benefits of earlier detection and reduced hospitalization. These trade-offs should be evaluated explicitly in health technology assessments.

## **9. Conclusion**

Uncertainty-aware deep learning has the potential to make long-term wearable cardiovascular monitoring safer and more clinically useful. By distinguishing between noise and genuine physiological anomalies, uncertainty estimates can reduce false alarms, detect silent failures, and support staged clinical escalation. However, these benefits will not emerge automatically from algorithmic improvements. They require a system-level perspective that integrates signal

quality processing, edge-cloud coordination, data governance, calibration, fairness, and clinical accountability. The architectural choices involved in uncertainty estimation have direct implications for energy use, latency, update frequency, and patient privacy. Data governance determines whether uncertainty signals can be interpreted correctly across diverse populations and changing deployment conditions. Policy and regulatory frameworks must evolve to accommodate continuous monitoring systems that learn over time. Future research should examine how uncertainty-aware models perform under realistic ambulatory conditions, how their outputs influence clinical workflows, and how governance structures can ensure that these systems do not widen existing health disparities. The long-term success of wearable cardiovascular monitoring will depend less on isolated accuracy gains than on the capacity of institutions to build trustworthy, equitable, and sustainable infrastructures around increasingly sophisticated models.

## References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
2. Hannun, A. Y., Rajpurkar, P., Haghpanahi, 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.
3. Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., Pellikka, P. A., & Friedman, P. A. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*, 25(1), 70–74.
4. Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P. S., & Meira, W. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nature Communications*, 11, 1760.
5. Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*.
6. Pantelopoulos, A., & Bourbakis, N. G. (2010). A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(1), 1–12.
7. 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.
8. Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6), 1236–1246.
9. 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.
10. 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.

11. 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.
12. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning*, 1321–1330.
13. Van Amersfoort, J., Smith, L., Teh, Y. W., & Gal, Y. (2020). Uncertainty estimation using a single deep deterministic neural network. *Proceedings of the 37th International Conference on Machine Learning*, 9690–9700.
14. 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.
15. 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.
16. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
17. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17, 195.
18. Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20, 310.
19. Saria, S., & Subbaswamy, A. (2019). Tutorial: Safe and reliable machine learning. *arXiv preprint arXiv:1904.07204*.
20. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.