

Transformer-Based Detection of Cardiac Arrhythmias from Ambulatory Photoplethysmography Signals

Claude Johansson

School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
johansson1979@uc.edu

Ravi C. Naidu

Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
ravinaidu850@colostate.edu

Maxime J. Fleming

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

Jerome M. Burton

Department of Computer Science, University of New Hampshire, Durham, NH, USA.
jeromeburton04@unh.edu

Abstract

Cardiac arrhythmias impose significant morbidity and mortality, with atrial fibrillation contributing to stroke, heart failure, and cognitive impairment. Ambulatory electrocardiography remains the reference standard for rhythm diagnosis, but prolonged use is constrained by electrode adhesion, patient burden, and cost. Photoplethysmography embedded in consumer wearables provides a lower-burden alternative for continuous cardiovascular monitoring. However, motion artifact, sensor heterogeneity, and algorithmic opacity limit clinical translation. This paper presents a system-level analysis of transformer-based detection of cardiac arrhythmias from ambulatory photoplethysmography signals. It examines sensing infrastructure, data governance, signal quality, architectural trade-offs, robustness, fairness, deployment, and policy implications. The discussion positions transformer models not only as sequence classifiers but as components within a broader sociotechnical monitoring pipeline requiring careful integration with preprocessing, explainability, regulatory oversight, and postmarket surveillance. The analysis emphasizes that performance improvements in offline datasets are insufficient for sustainable clinical utility. Attention to energy consumption, data drift, demographic equity, and governance is equally important. The paper synthesizes developments in time-series attention, motion robustness, and responsible machine learning to outline a forward-looking framework for next-generation ambulatory cardiac monitoring systems.

Keywords

transformer models; photoplethysmography; arrhythmia detection; ambulatory monitoring; health informatics; algorithmic fairness.

1. Introduction

Cardiac arrhythmias remain a principal contributor to global cardiovascular disability, with atrial fibrillation increasing risk for ischemic stroke, heart failure, and cognitive decline [1]. The clinical need for early and continuous rhythm assessment has motivated a shift from intermittent clinic-based electrocardiography toward ambulatory and consumer-oriented monitoring [2]. Deep neural networks have shown capacity to identify rhythm disorders from single-lead electrocardiograms with performance approaching expert cardiologists, yet the acquisition of electrocardiographic signals over long periods remains burdensome because of electrodes, skin preparation, and patient adherence [2]. Photoplethysmography offers an alternative by detecting blood volume changes through optical sensors, enabling pulse rate and rhythm analysis in devices that are already worn for daily activity tracking [3]. A growing body of work has demonstrated that photoplethysmographic signals can support atrial fibrillation screening and arrhythmia classification outside specialized cardiology environments [4]. Large-scale smartwatch studies have shown that consumer photoplethysmography can identify irregular pulses with clinically actionable findings, although confirmatory electrocardiography is still required [5]. Mobile photoplethysmographic screening has also been evaluated in broad populations, confirming feasibility and underlining the need for high-specificity triage [6]. Passive monitoring approaches have further shifted the burden from active user engagement to background data collection, creating both opportunities and governance challenges [7]. Transformers, originally developed for natural language processing, have emerged as general-purpose sequence and representation learners capable of capturing long-range dependencies [8]. This paper examines their application to arrhythmia detection from ambulatory photoplethysmography signals from a systems perspective rather than focusing solely on model architecture or benchmark accuracy.

2. Clinical Context and Sensing Infrastructure

The clinical utility of photoplethysmography-based arrhythmia detection depends on how well the sensing infrastructure matches the temporal and physiological characteristics of arrhythmias. Paroxysmal events may be brief and infrequent, so monitoring must be continuous, comfortable, and energy efficient. Electrocardiographic patches provide high diagnostic information but are limited by wear time and patient adherence. Photoplethysmography sensors in wrist-worn devices collect pulse information continuously but introduce challenges related to sensor positioning, skin tone, ambient light, and movement. These trade-offs are structural rather than purely algorithmic. A monitoring system must balance diagnostic yield against false alarms, user burden, and the cost of confirmatory testing. Transformer models can process long sequences and attend to global rhythm patterns, making them conceptually suitable for ambulatory data with irregular temporal spacing. Pretrained sequence models have shown that representations learned from large unlabeled corpora can transfer to downstream tasks, a principle relevant to physiological time series [9]. Time-series transformers have expanded this idea to multivariate sensor streams by learning embeddings that preserve temporal structure and enable downstream classification and regression [10]. Such approaches may reduce the need for hand-crafted photoplethysmographic features while still requiring careful preprocessing and domain adaptation.

From an infrastructure standpoint, the choice of model family interacts with data storage, labeling, and transmission. Ambulatory photoplethysmography monitoring generates large volumes of continuous data, much of which is normal sinus rhythm. Cloud-based training may permit large models, but edge deployment imposes memory and latency constraints.

Transfer learning strategies can support model compression and adaptation to heterogeneous devices. The representation capacity of transformers may be particularly valuable when sensor streams include accelerometer, heart rate, and photoplethysmography signals with varying sampling rates. However, these advantages do not eliminate the need for robust signal quality management and governance, because attention mechanisms can amplify spurious correlations if motion artifacts are not addressed. The next section examines these signal quality considerations.

3. Signal Quality and Motion Robustness

Motion artifacts are among the most consequential threats to photoplethysmography-based arrhythmia detection. Physical activity induces variations in capillary blood flow and sensor contact that can mimic irregular pulses or obscure existing rhythm abnormalities. Unless these artifacts are removed before model ingestion, transformer-based classifiers may learn to associate movement-related waveform distortions with clinical labels, creating brittle models that fail outside controlled conditions. Adaptive notch filtering has been shown to recover photoplethysmographic signals during physical activity, illustrating the importance of preprocessing stages that are explicitly designed for ambulatory contexts [11]. Such signal recovery methods operate before deep learning and can improve the stability of downstream representations. The integration of signal processing and learning should be understood as a layered control problem in which each layer reduces uncertainty before the next layer attempts classification.

The placement of signal quality controls within a monitoring pipeline has structural implications. If quality assessment occurs only at the cloud after transmission, bandwidth and energy are wasted on corrupted segments. If it occurs at the edge, false rejection or acceptance can alter the data distribution seen by the model. A robust system may use adaptive filtering as part of a preprocessing subsystem that also tracks accelerometer context, sensor contact, and heart rate plausibility. This preprocessing layer is not independent from the transformer architecture because input normalization and artifact management change the statistical properties that attention heads learn. Governance requirements further demand that signal rejection decisions be auditable; otherwise, missing data can disproportionately affect populations whose activity patterns or sensor fit differ from development cohorts. These interactions show that motion robustness is not merely a technical preprocessing issue but a sociotechnical design constraint.

4. Transformer-Based Modeling Architecture

Before the adoption of transformer architectures, recurrent neural networks dominated physiological time-series modeling because of their ability to maintain hidden states across time steps. Long short-term memory networks introduced gating mechanisms to address vanishing gradients and have been widely used for sequential health data [12]. Gated recurrent units simplified these mechanisms and reduced computational cost while retaining memory for many tasks [13]. Recurrent models nevertheless process time steps sequentially, which can limit parallelization and make learning long-range dependencies difficult. Transformer architectures replace recurrence with self-attention, allowing each position to attend to all other positions in the sequence. This design is attractive for ambulatory photoplethysmography because arrhythmic patterns may depend on distant context, such as a preceding exercise episode or gradual changes in pulse morphology. The vision transformer further demonstrated that images can be treated as sequences of patches, suggesting that physiological waveforms can similarly be divided into segments and encoded as tokens [14].

This tokenization perspective supports flexible fusion of multiple sensor channels and may improve generalization across devices.

Architectural choices are embedded within broader data quality and evaluation contexts. Photoplethysmographic morphology varies across sensor locations, skin types, and device designs, so tokenization should be informed by the physiological basis of pulse waveforms rather than purely empirical patch splitting [15]. The transformer encoder can then learn representations that are less dependent on local waveform amplitude and more attentive to temporal regularity, phase relationships, and interbeat intervals. However, self-attention over long recordings may require substantial memory and compute; therefore, patch size and sequence length are structural parameters that control the trade-off between global context and resource demand. In ambulatory settings, where devices process hours or days of data, these choices influence battery life, heat dissipation, and the feasibility of continuous background inference. Model architectures that adapt attention span or use hierarchical token aggregation may be preferable to full-sequence attention for wearable deployment. Cross-channel attention between photoplethysmography and accelerometer streams can also reduce motion-induced false positives by conditioning classification on activity context. Such multimodal designs are closer to clinical reasoning because clinicians interpret pulse findings in relation to patient movement and reported symptoms.

Furthermore, pretraining on large unlabeled physiological datasets can provide transferable representations that reduce the need for extensive labeled arrhythmia data. Yet pretraining introduces additional questions about dataset composition, privacy, and representation bias. A transformer trained on heterogeneous signals from healthy volunteers may not generalize to patients with low ejection fraction, frequent ectopy, or poor peripheral perfusion. Consequently, transfer learning must be accompanied by domain adaptation and subgroup-specific evaluation. The architectural separation between tokenization, attention, and classification can simplify modular auditing, but it can also obscure how intermediate representations influence final decisions. These limitations motivate a broader system-level analysis that treats the model as one component of a larger monitoring infrastructure.

5. System-Level Trade-offs and Resource Constraints

System-level deployment involves a complex set of trade-offs among accuracy, latency, energy, storage, privacy, and update frequency. The quality of photoplethysmography inputs is not constant across time or across individuals; therefore, a signal quality index should be estimated before inference to avoid feeding noisy segments into the classifier. Orphanidou and colleagues [16] developed signal quality indices for electrocardiogram and photoplethysmogram signals that can inform gating decisions in wearable pipelines. When signal quality is low, the system may reject the segment, request a new measurement, or trigger a lower-confidence notification rather than silently producing an unreliable classification. This gating layer has direct consequences for false negatives and false positives, both of which carry clinical and operational costs. A system that rejects too aggressively may miss important paroxysmal events, while one that accepts corrupted data may overwhelm users with false alarms and reduce trust.

Deploying transformer models at the edge requires attention to energy consumption because self-attention operations scale quadratically with sequence length and can drain batteries if not optimized. Strubell and colleagues [17] documented the environmental impact of deep learning in natural language processing, but similar principles apply to physiological monitoring. Although a single inference may consume little energy, continuous monitoring

over weeks or months aggregates into significant resource use. Distillation, quantization, and pruning can reduce model size and latency; however, compression must be validated for subtle arrhythmia patterns, because overaggressive compression may eliminate rare but clinically important features. The structural decision to place computation on the device, in a gateway, or in the cloud alters privacy, reliability, and battery life. Federated or distributed learning can reduce raw data transmission and support personalization without centralizing sensitive physiological data; however, this introduces synchronization, fairness, and governance complexities [18]. For instance, federated updates from devices with biased sensor characteristics may degrade global model performance if aggregation strategies do not account for population heterogeneity.

6. Data Governance, Fairness, and Regulatory Implications

Data governance for ambulatory photoplethysmography requires balancing clinical benefit with privacy, consent, and data provenance. The widespread collection of continuous physiological data creates risks around reidentification, insurance discrimination, and unregulated secondary uses. Vayena and colleagues [19] argue that machine learning in medicine requires governance that extends beyond technical accuracy to include transparency, accountability, and public trust. Documentation mechanisms such as model cards can make operational characteristics and limitations explicit, enabling regulators and clinicians to assess fitness for purpose [20]. These artifacts are not merely administrative; they structure the organizational memory of development choices and support postmarket audits. Without such documentation, institutions cannot reliably know when a model is drifting, which subgroups were underrepresented, or what thresholds were used for clinical action.

Fairness is not guaranteed by large datasets. Obermeyer and colleagues [21] showed that health algorithms can reproduce systemic biases even when protected attributes are absent. For photoplethysmography, skin tone, perfusion, and sensor fit can interact with optical signal quality, and a model may underperform for darker skin if development data are not representative. This is particularly relevant because pulse oximetry and photoplethysmography have documented performance variability across skin pigmentation, and arrhythmia detection must not inherit those limitations. The regulatory pathway must therefore require performance evaluation across demographic and clinical subgroups, including patients with low perfusion, high motion, and coexisting conditions like atrial ectopy or conduction disorders. Benchmarks such as the PhysioNet/Computing in Cardiology Challenge 2017 provide useful but limited comparisons, because short recordings and curated labels do not capture the noise and distribution shifts of long-term ambulatory monitoring [22].

7. Deployment, Postmarket Surveillance, and Sustainability

Deployment of transformer-based arrhythmia detection must extend beyond model accuracy to include sensor drift, firmware updates, user interface design, and clinician workflow integration. Postmarket surveillance is essential because ambulatory populations differ from clinical development cohorts, and new arrhythmia patterns may emerge. Benjamins and colleagues [23] reviewed the increasing number of artificial intelligence-based medical devices approved by regulators, highlighting the need for ongoing performance monitoring rather than one-time validation. Continuous data collection enables learning health systems, but it also raises questions about who monitors model drift and how updates are authorized. Wearable photoplethysmography devices can be validated in controlled settings, yet field performance during daily activities may diverge substantially from laboratory results [24]. A

transformer model trained on short, motion-controlled segments may encounter unexpected artifact patterns when deployed at scale, so robust streaming evaluation and human oversight are required.

Sustainability considerations include hardware life cycles, e-waste, energy use, and the carbon footprint of both training and inference. A transformer-based system may have higher computational demands than simpler classifiers, so deployment must justify that cost with improved patient outcomes or reduced overdiagnosis. The integration of explainable attention maps into clinical dashboards can help users understand why a notification was generated, but these maps must be interpreted cautiously because attention weights do not guarantee causal attribution [25]. Interpretability must be supplemented with rigorous subgroup performance, uncertainty estimation, and structured clinician feedback. Without such supports, attention visualizations may create an illusion of understanding that obscures hidden failure modes.

8. Limitations and Future Perspectives

The evidence base for transformer-based photoplethysmographic arrhythmia detection remains dominated by retrospective and single-site studies. Prospective trials with clinically adjudicated outcomes are needed to assess whether such models reduce stroke, hospitalization, or mortality. The generalizability of attention-based representations across device generations, skin types, and patient comorbidities has not been established. Future research should combine adaptive filtering, signal quality indices, and multimodal attention with decentralized training frameworks that preserve privacy and permit continuous learning. Evaluation should include not only aggregate area under the curve but also calibration, false alarm burden, and time-to-detection for paroxysmal events. These metrics are more closely aligned with clinical utility than static benchmark accuracy. Moreover, equitable access to wearable monitoring and downstream confirmatory care must be addressed, otherwise algorithmic advances may widen existing health disparities.

9. Conclusion

Transformer-based detection of cardiac arrhythmias from ambulatory photoplethysmography signals represents a promising convergence of sequence modeling and pervasive sensing. This paper has argued that clinical value cannot be reduced to architectural novelty or offline performance. The integration of photoplethysmography, motion robustness, attention mechanisms, governance, fairness, and postmarket monitoring forms a sociotechnical system that requires coordinated design across signal processing, model engineering, regulation, and clinical workflow. Adaptive filtering methods such as the one described in [11] illustrate the foundational role of signal preprocessing, while attention-based architectures offer powerful representational capacity if their resource and interpretability limitations are managed. Future work should prioritize prospective validation, representative training data, energy-efficient deployment, and governance structures that protect patients while enabling responsible innovation.

References

1. Benjamin, E. J., Muntner, P., Alonso, A., Bittencourt, M. S., Callaway, C. W., Carson, A. P., Chamberlain, A. M., Chang, A. R., Cheng, S., Das, S. R., Delling, F. N., Djousse, L., Elkind, M. S. V., Ferguson, J. F., Fornage, M., Jordan, L. C., Khan, S. S., Kissela, B. M., Knutson, K. L., ... Virani, S. S. (2019). Heart disease and stroke statistics—2019 update: A report from the American Heart Association. *Circulation*, 139(10), e56–e528.

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. 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.
4. Poh, M. Z., Poh, Y. C., Chan, P. H., Wong, C. K., Pun, L., Leung, W. W., Wong, M. F., Chu, M. M., Leung, P. Y., Chan, C. K., Chan, K. P., Lau, C. P., & Siu, C. W. (2018). Diagnostic assessment of a deep learning system for detecting atrial fibrillation using photoplethysmography. *Nature Medicine*, 24(8), 1331–1337.
5. 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., ... Turakhia, M. P. (2019). Large-scale assessment of a smartwatch to identify atrial fibrillation. *New England Journal of Medicine*, 381(20), 1909–1917.
6. Guo, Y., Wang, H., Zhang, H., Liu, T., Liang, Z., Xia, Y., Yan, L., Xing, Y., Shi, H., Li, S., Liu, Y., Liu, F., Feng, M., Chen, Y., & Lip, G. Y. H. (2019). Mobile photoplethysmographic technology to detect atrial fibrillation. *Journal of the American College of Cardiology*, 74(19), 2365–2375.
7. Tison, G. H., Sanchez, J. M., Ballinger, B., Singh, A., Olgin, J. E., Pletcher, M. J., Vittinghoff, E., Lee, E. S., Fan, S. M., Gladstone, R. A., Mikell, C., Sohoni, N., Hsieh, J., & Marcus, G. M. (2018). Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiology*, 3(5), 409–416.
8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
9. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171–4186.
10. Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. (2021). A transformer-based framework for multivariate time series representation learning. *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 2114–2124.
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. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
13. Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for

statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 1724–1734.

14. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations.
15. Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. *Physiological Measurement*, 28(3), R1–R39.
16. Orphanidou, C., Bonnici, T., Charlton, P., Clifton, D., Vallance, D., & Tarassenko, L. (2015). Signal quality indices for the electrocardiogram and photoplethysmogram: Derivation and applications in ambulatory monitoring. *IEEE Journal of Biomedical and Health Informatics*, 19(3), 832–838.
17. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645–3650.
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.
19. Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine*, 15(11), e1002689.
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. 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.
22. Clifford, G. D., Liu, C., Moody, B., Lehman, L. H., Silva, I., Li, Q., Johnson, A. E. W., & Mark, R. G. (2017). AF classification from a short single lead ECG recording: The PhysioNet/Computing in Cardiology Challenge 2017. *Computing in Cardiology*, 44, 1–4.
23. Benjamins, S., Dhunoo, P., & Mesko, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database. *npj Digital Medicine*, 3, 118.
24. Bonomi, A. G., Schipper, F., Eerikäinen, L. M., Margarito, J., van Dinther, R., Muesch, G., de Morree, H. M., Aarts, R. M., Babaeizadeh, S., McManus, D. D., & Dekker, L. R. C. (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.
25. Jain, S., & Wallace, B. C. (2019). Attention is not explanation. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 3543–3556.