ORIGINAL RESEARCH Open access

Edge AI on Smartwatches for Atrial Fibrillation Detection: A Perspective on Real-Time Processing, Power Efficiency, and Clinical Integration

Bruno Martins[1*], Lucas Pereira[1], Renata Azevedo[2], Pedro Costa[1]

Abstract

Atrial fibrillation (AFib) is a major and often undiagnosed risk factor for ischemic stroke, with paroxysmal episodes that frequently evade conventional intermittent monitoring. Wearable devices combining photoplethysmography (PPG) and single-lead ECG have enabled large-scale AFib screening, but many current systems rely on cloud-based processing, introducing latency, connectivity dependence, and privacy concerns. While clinical studies demonstrate promising detection performance, real-world deployment remains limited by the lack of fully continuous, autonomous operation. Edge artificial intelligence (AI), which enables on-device deep-learning inference directly on smartwatches, represents a key advancement toward real-time, scalable AFib detection. By eliminating reliance on cloud infrastructure, edge AI reduces latency, enhances privacy, and supports immediate alerts during transient arrhythmic events. However, practical implementation requires careful optimization of model efficiency, power consumption, and hardware constraints alongside clinical validation. Future progress will depend on multi-objective design strategies that integrate accuracy, latency, and energy efficiency, as well as collaboration among engineers, clinicians, and regulators. Addressing challenges such as alert fatigue, equitable access, and data governance will be essential. Ultimately, edge AI has the potential to transform AFib management from reactive diagnosis to continuous, preventive monitoring, functioning as an unobtrusive, always-available cardiac safeguard.

Keywords Edge AI, Smartwatch, Atrial fibrillation, PPG, Real-time processing, Power efficiency

*Correspondence:
Bruno Martins
bruno.martins@gmail.com
[1] Department of Healthcare Data Modeling, Faculty of Engineering, University of Minho, Braga, Portugal
[2] Department of Clinical AI Systems, Faculty of Medicine, University of Porto, Porto, Portugal

Introduction

Atrial fibrillation (AFib) is the most common cardiac arrhythmia, affecting 1-2% of the general population and up to 10% of those over 80 [1, 2]. Undiagnosed AFib increases stroke risk five-fold, yet timely anticoagulation can dramatically mitigate this danger when detection occurs early [3]. Large-scale digital health initiatives have begun to quantify the hidden burden of silent AFib, highlighting the limitations of traditional Holter or event monitors for long-term surveillance [4]. As consumer wearables proliferate, the opportunity to embed intelligent screening directly into daily life has never been greater [5].

Smartwatches from leading manufacturers now incorporate AFib detection features based on PPG and ECG signals [6, 7]. Many implementations still transmit raw or lightly processed data to cloud servers for analysis, which creates latency, privacy vulnerabilities, and accelerated battery drain [8]. Early studies have validated these hybrid approaches in controlled cohorts, yet continuous real-world use exposes the fragility of network-dependent pipelines [9]. The engineering community must therefore confront the inherent constraints of battery-powered, wrist-worn platforms [10].

Edge AI—executing sophisticated deep-learning models directly on the smartwatch—offers a compelling alternative to cloud-centric designs [11, 12]. On-device inference enables immediate rhythm classification without constant data transmission, preserving battery life and user privacy while delivering real-time alerts [13]. Preliminary explorations of lightweight neural networks on wearables already hint at feasibility, yet substantial optimization remains necessary before widespread clinical adoption [14, 15]. This architectural pivot demands fresh thinking about model design, hardware co-design, and clinical workflow alignment [16].

This Perspective examines the transition to on-device edge AI for smartwatch-based AFib detection. We analyze the core constraints of real-time processing and power efficiency, evaluate emerging engineering strategies such as model compression and hardware acceleration, and interrogate the challenges of clinical integration [17]. We argue that multi-objective optimization—balancing accuracy, latency, and power—is the central design problem confronting the field [18]. The following sections chart a roadmap for realizing clinically viable edge AI on resource-constrained wearables [19].

AFib Detection Landscape

Clinical significance and screening paradigms

AFib confers a five-fold elevated risk of stroke, yet up to one-third of cases remain undetected until a cerebrovascular event occurs [1, 3]. Paroxysmal episodes are particularly elusive because they may last only minutes and evade periodic clinic-based ECGs, underscoring the value of opportunistic or systematic screening strategies [2]. Digital health studies have shown that consumer wearables can surface otherwise silent AFib, potentially enabling earlier anticoagulation and stroke prevention [4, 20]. Future screening paradigms must therefore evolve toward continuous, passive monitoring that integrates seamlessly into everyday life without overburdening healthcare systems [21].

The distinction between opportunistic and systematic screening carries profound implications for public health resource allocation [1]. While systematic population-level programs risk overwhelming clinics with false positives, opportunistic use during routine smartwatch interactions offers a low-friction entry point [3]. Longitudinal data from large cohorts reveal that even brief detected episodes correlate with increased thromboembolic risk, reinforcing the need for sensitive yet specific algorithms [5]. As edge AI matures, these devices could dynamically adjust screening intensity based on individual risk profiles, moving the field closer to personalized preventive cardiology [6].

Current smartwatch AFib detection

Contemporary smartwatches employ PPG-based algorithms or intermittent single-lead ECG recordings to flag possible AFib episodes [3, 7]. Landmark prospective studies have established proof-of-concept for consumer-grade detection at scale, reporting encouraging sensitivity and specificity in real-world ambulatory settings [8, 9]. Nevertheless, reliance on cloud processing introduces unavoidable delays between signal acquisition and clinical notification, limiting utility during transient arrhythmias [10]. These hybrid systems also expose users to connectivity failures and heightened privacy concerns inherent in continuous data uploads [11].

Single-lead ECG implementations on newer devices have improved specificity compared with PPG-only approaches, yet both modalities remain susceptible to motion artifacts and skin-tone variability [12]. Validation efforts continue to refine positive predictive values, particularly in older adults and those with comorbidities [13]. Looking forward, the integration of edge AI promises to retain the strengths of existing hardware while eliminating the latency and transmission overhead that currently constrain performance [14]. The next generation of smartwatch AFib detectors must therefore prioritize on-device intelligence to bridge the gap between consumer convenience and clinical reliability [15].

Edge AI on Smartwatches

What edge AI enables

Edge AI empowers smartwatches to classify cardiac rhythms in real time without ever leaving the device, delivering instantaneous alerts that could prompt life-saving behavioral or therapeutic responses [16]. By keeping raw PPG and ECG waveforms local, this paradigm safeguards patient privacy and eliminates recurring data-transmission costs that erode battery life [17]. Early prototypes of on-device neural networks already demonstrate the capacity to run continuous inference within the tight thermal and power budgets of wrist-worn hardware [18]. As these capabilities mature, clinicians will gain access to richer, context-aware rhythm data that cloud-dependent systems simply cannot provide [19].

The shift to fully on-device processing also unlocks new clinical use cases, such as adaptive monitoring that escalates fidelity only when preliminary low-power detectors flag anomalies [22]. Reduced latency means alerts can reach users within seconds of arrhythmia onset rather than minutes or hours later [23]. This immediacy is especially valuable for paroxysmal AFib, where timely notification might prevent stroke or heart-failure exacerbation [24]. Ultimately, edge AI transforms the smartwatch from a passive data collector into an active clinical companion embedded in the patient's daily routine [25].

Figure 1 illustrates the hierarchical edge AI pipeline enabling real-time, power-efficient AFib detection on smartwatches through cascaded on-device inference and selective clinical data transmission.

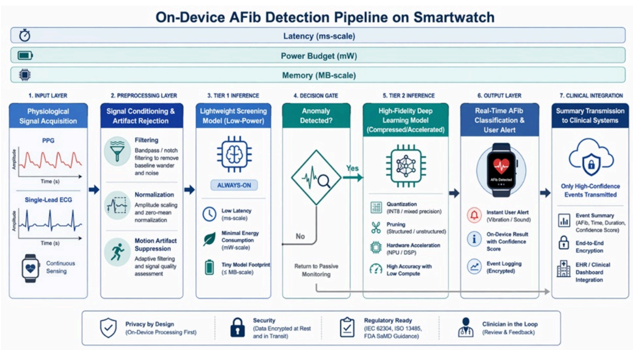

Figure 1. Hierarchical Edge AI Architecture for Real-Time Atrial Fibrillation Detection on Smartwatches

Hardware constraints

Smartwatch processors operate under severe limitations: memory measured in megabytes rather than gigabytes, CPU/GPU cycles shared with multiple always-on sensors, and battery capacities that must sustain multi-day usage [26]. Thermal throttling further restricts sustained high-performance inference, forcing designers to trade computational depth for acceptable skin temperatures [27]. Existing neural processing units and digital signal processors offer partial acceleration, yet they remain underutilized in current AFib pipelines that still offload heavy lifting to the cloud [10]. Future edge AI solutions must therefore be co-designed with these exact hardware realities rather than retrofitted afterward [11].

Memory footprints, power envelopes, and interrupt latency collectively dictate which model architectures can even be deployed [12]. Quantized and pruned networks that once seemed exotic are rapidly becoming mandatory for viable on-wrist deployment [16, 17]. As semiconductor roadmaps deliver more efficient neural engines tailored for wearables, the performance ceiling will rise, yet the fundamental constraints of size and energy will persist [18]. Engineers must therefore treat hardware limitations as first-class design variables rather than obstacles to be worked around [19].

Why not just use the cloud?

Cloud inference, while accurate, imposes unavoidable network round-trip delays that render real-time AFib alerts clinically suboptimal during fleeting episodes [22]. Continuous data transmission also accelerates battery depletion far beyond what on-device processing requires, shortening device uptime and degrading user experience [23]. Dependence on cellular or Wi-Fi connectivity further excludes users in remote or low-coverage areas, limiting equitable access to advanced monitoring [24]. Edge AI sidesteps these drawbacks by localizing computation where the signal originates [25].

Privacy regulations increasingly scrutinize the transmission of identifiable physiological data, making on-device inference an attractive compliance strategy [26]. Cloud pipelines also incur recurring infrastructure costs that scale poorly with millions of always-connected wearables [1]. By contrast, edge AI shifts the computational burden to the device itself, reducing long-term operational expenses for manufacturers and health systems alike [3]. The combination of latency reduction, privacy preservation, and cost efficiency positions edge AI as the logical evolution beyond today's hybrid architectures [4].

Real-time Processing Constraints

Continuous vs on-demand monitoring

Continuous monitoring demands that AFib detection models remain active at all times, placing unrelenting pressure on both processor availability and battery reserves [5]. On-demand or user-triggered modes conserve energy but risk missing asymptomatic episodes that constitute the majority of undiagnosed AFib burden [6]. Periodic sampling offers a middle ground yet still introduces blind intervals during which silent arrhythmias can evade detection entirely [27]. Future edge AI systems must therefore intelligently toggle between modes based on contextual cues such as heart-rate variability or user activity [7].

The choice between always-on and event-driven inference directly influences clinical sensitivity and device acceptability [8]. Always-on approaches maximize detection opportunities yet accelerate battery drain and heighten alert fatigue when false positives accumulate [9]. Adaptive architectures that begin with lightweight screening filters before engaging heavier models strike a promising balance [10]. As edge hardware improves, the field will increasingly favor continuous yet power-aware monitoring that feels invisible to the wearer while remaining clinically vigilant [11].

Latency requirements for clinical utility

AFib episodes frequently last only minutes, necessitating detection latency measured in seconds rather than minutes if alerts are to influence immediate patient behavior or trigger downstream care pathways [12]. Near-real-time cloud processing often exceeds clinically actionable windows, especially during exercise or travel when connectivity fluctuates [13]. Edge AI collapses this latency to the millisecond scale of on-chip inference, enabling proactive notifications before the episode self-terminates [14]. This temporal advantage could prove decisive in preventing thromboembolic complications [15].

Latency targets must also account for downstream clinical workflows, including user acknowledgment and confirmatory ECG acquisition [16]. Excessive delay between detection and alert undermines the very purpose of continuous monitoring [17]. Conversely, sub-second inference that floods users with unfiltered notifications risks desensitization and alert fatigue [18]. The optimal latency therefore emerges from careful co-design of model speed, notification logic, and human-factors engineering tailored to real-world ambulatory settings [19].

Window size and overlap

Sliding-window strategies balance temporal resolution against computational cost: longer windows improve rhythm classification accuracy yet increase per-inference latency and energy draw [22]. Overly short windows risk fragmenting AFib episodes into ambiguous segments, while excessive overlap inflates redundant computation on resource-limited hardware [23]. Edge AI designers must therefore optimize window parameters jointly with model architecture to maintain clinical sensitivity without sacrificing battery life [24]. Adaptive window sizing that expands during suspected arrhythmic periods offers an elegant compromise [25].

Overlap ratios further modulate the trade-off between detection reliability and processing overhead [26]. High-overlap regimes ensure no episode is missed yet multiply floating-point operations per second, challenging the thermal envelope of plastic-cased wearables [27]. Future systems may employ hierarchical windowing—coarse low-power scans followed by fine-grained high-accuracy analysis only when warranted [7]. Such cascaded designs exemplify the multi-objective thinking required to make real-time edge AI viable for long-term AFib surveillance [8].

Power Efficiency Strategies

Model compression techniques

Quantization from 32-bit to 8-bit representations, combined with structured pruning, can shrink deep neural networks for AFib detection to fractions of their original size while preserving acceptable accuracy on PPG and ECG signals [16, 17]. Knowledge distillation transfers capabilities from large teacher models to compact student networks specifically tuned for wearable constraints [18]. Neural architecture search tailored for tinyML further automates the discovery of power-efficient topologies that fit within smartwatch memory limits [19]. These compression pipelines are no longer optional but foundational to any deployable edge AI solution [22].

Post-training quantization and sparsity-aware training have already enabled real-time inference on low-power microcontrollers, yet AFib-specific models require careful calibration to avoid degrading sensitivity in noisy ambulatory data [23]. Iterative pruning guided by hardware-in-the-loop feedback ensures that accuracy degradation remains clinically insignificant [24]. As these techniques mature, the field will witness models that deliver near-cloud performance at a fraction of the energy cost [25]. The

https://doi.org/10.51847/u591867573

coming generation of wearable AFib detectors will therefore owe much of their practicality to advances in systematic model compression [26].

Hardware acceleration

Dedicated neural processing units and digital signal processors embedded in modern smartwatch chipsets accelerate matrix operations far more efficiently than general-purpose CPUs [10, 11]. Platforms featuring specialized accelerators for convolutional and recurrent layers can sustain continuous AFib inference within the milliwatt power budgets required for all-day wear [12]. Co-design between algorithm developers and silicon architects will further tailor these accelerators to the unique arithmetic patterns of cardiac time-series data [13]. The result is inference that feels instantaneous yet consumes orders of magnitude less energy than software-only implementations [14].

Emerging low-power AI engines already demonstrate the feasibility of running quantized arrhythmia models entirely on-device without perceptible impact on battery longevity [15]. Thermal management remains critical, however, because even efficient accelerators generate localized heat that users notice on the wrist [16]. Future hardware roadmaps must therefore prioritize not only peak throughput but also energy-per-inference metrics benchmarked under realistic wearable workloads [17]. When hardware acceleration and model compression advance in lockstep, edge AI for AFib detection will transition from laboratory curiosity to everyday clinical tool [18].

Table 1 provides a conceptual mapping between key edge AI strategies and their downstream clinical and system-level impacts.

Table 1. Analytical Mapping of Edge AI Design Strategies to Clinical and System-Level Outcomes

Edge AI Strategy	System-Level Effect	Clinical Benefit	Impl… De…
Model Compression (Quantization, Pruning)	Reduced memory footprint and energy per inference	Enables continuous monitoring without battery compromise	Hard… t… p…
Hardware Acceleration (NPU/DSP)	Orders-of-magnitude efficiency in matrix operations	Near-instantaneous AFib detection	Co-… ar…
Duty Cycling & Adaptive Inference	Dynamic allocation of computational resources	Maintains sensitivity while reducing unnecessary processing	Con… co…
On-Device Personalization	User-specific model adaptation	Reduced false positives/negatives across diverse populations	E… inc… I… r…
Multimodal Sensor Fusion	Improved signal robustness and context awareness	Enhanced discrimination between true arrhythmia and artifacts	Sync… acro…
Edge-Only Processing (No Cloud Dependency)	Elimination of network latency and data transmission	Immediate alerts during transient AFib episodes	Fully… er…
Selective Data Transmission	Reduced bandwidth and privacy exposure	Streamlined clinical workflows	Cc… sc… sum…

Duty cycling and adaptive inference

Duty-cycling strategies—running lightweight screening models at high frequency while reserving full deep-learning inference for suspected events—dramatically extend battery life without compromising detection sensitivity [19]. Adaptive inference further refines this approach by modulating model complexity according to real-time context such as motion intensity or recent heart-rate trends [22]. Such hierarchical pipelines ensure that the power-hungry components activate only when clinically justified, aligning computational effort with actual arrhythmic risk [23]. The outcome is a smartwatch that remains vigilant around the clock yet feels indistinguishable from an ordinary fitness tracker [24].

Context-aware triggering also mitigates alert fatigue by suppressing unnecessary high-fidelity analyses during periods of stable sinus rhythm [25]. On-device reinforcement learning or simple rule-based controllers can learn user-specific patterns over time, further optimizing duty cycles for individual physiology [26]. As these adaptive mechanisms mature, power efficiency will cease to be a limiting factor and instead become a tunable parameter that clinicians and patients adjust together [27]. Edge AI, paired with intelligent duty cycling, therefore promises to deliver continuous AFib monitoring that is both clinically robust and sustainably powered [7].

Clinical Integration Challenges

False positives and alert fatigue

False positives arising from motion artifacts during exercise, hand washing, or daily activities continue to undermine user trust in wearable AFib detection systems [8]. Early large-scale validations of PPG-based algorithms revealed that real-world noise can inflate false-alert rates, leading to alert fatigue that discourages both patients and clinicians from acting on notifications [9]. Edge AI offers a path forward by embedding lightweight artifact-rejection layers directly on the device, dynamically suppressing unreliable signals before they trigger user-facing alerts [10]. As these on-device filters improve, the balance between sensitivity and tolerability will determine whether continuous monitoring becomes a sustainable clinical asset rather than a source of distraction [11].

Clinicians already grapple with the downstream burden of excessive notifications that flood electronic health records and outpatient schedules [12]. Prospective studies of consumer wearables have shown that unfiltered alerts can overwhelm primary-care workflows, delaying genuine cases amid the noise [13]. Future edge AI architectures must therefore incorporate user-specific thresholding and contextual awareness to minimize fatigue while preserving detection of clinically significant episodes [14]. Only through such intelligent on-device curation can wearable AFib screening achieve the seamless integration required for long-term population health impact [15].

Clinical workflow integration

Integrating smartwatch-derived AFib alerts into existing clinical pathways demands more than technical accuracy; it requires clear escalation protocols from device notification to confirmatory testing and anticoagulation decisions [16]. Current hybrid systems often leave gaps between consumer alerts and physician action, resulting in delayed follow-up or unnecessary cardiology referrals [17]. Edge AI can streamline this cascade by transmitting only high-confidence summaries rather than raw waveforms, thereby reducing data overload for electronic health record systems [18]. The coming decade will test whether such streamlined workflows can translate wearable signals into measurable reductions in stroke incidence [19].

Real-world deployment further highlights the need for shared decision-making tools that help patients and providers interpret edge-detected episodes in context [22]. Validation cohorts have demonstrated that without standardized handoff mechanisms, many flagged events never reach formal ECG confirmation [23]. Forward-looking clinical integration must therefore embed decision-support algorithms at the device level, guiding users toward appropriate next steps without requiring constant physician oversight [24]. When edge AI and clinical workflows co-evolve, wearable AFib detection will evolve from an isolated consumer feature into a true extension of preventive cardiology [25].

Trade-Offs and Design Decisions

Accuracy vs power

Larger, more accurate deep-learning models for AFib detection inevitably increase power draw, threatening the multi-day battery life that consumers expect from smartwatches [26]. Hardware-constrained platforms force engineers to sacrifice some predictive performance for sustainable operation, creating a fundamental tension that cloud-based systems never encounter [27]. Emerging compression pipelines have already narrowed this gap, yet the optimal operating point remains use-case dependent—screening applications may tolerate slightly lower accuracy to preserve battery, while diagnostic-grade monitoring cannot [7]. The field must therefore treat power as a first-class clinical metric rather than an engineering footnote [8].

Multi-objective optimization frameworks will become essential as edge AI matures, allowing designers to navigate the Pareto front of accuracy and energy explicitly [9]. Preliminary explorations of quantized and pruned networks on wearable hardware confirm that clinically acceptable performance can coexist with extended battery endurance [10]. Future smartwatch platforms equipped with more efficient neural engines will expand the feasible design space, yet the core trade-off will persist [11]. Ultimately, the winning architectures will be those that deliver “good enough” accuracy at truly wearable power levels rather than chasing marginal gains at unsustainable cost [12].

Table 2 delineates the fundamental multi-objective trade-offs that govern the design of edge AI systems for smartwatch-based AFib detection.

Table 2. Multi-Objective Trade-off Landscape in Edge AI–Enabled Smartwatch AFib Detection

Design Dimension	Competing Objectives	Engineering Strategies	Clinical I
Accuracy vs Power	High diagnostic precision vs multi-day battery life	Quantization, pruning, knowledge distillation, tinyML architectures	Sustained r clinically ser
Latency vs Specificity	Immediate alerts vs reduced false positives	Cascaded inference, threshold tuning, adaptive filtering	Timely inte transient A
Continuous vs On-Demand Monitoring	Maximal detection coverage vs energy conservation	Duty cycling, event-triggered inference, adaptive sampling	Dete asymptoma A
Model Complexity vs Hardware Constraints	Deep architectures vs limited memory/compute	Hardware-aware neural architecture search, accelerator utilization	Feasibility dep
Sensitivity vs Clinical Workflow Burden	High case detection vs manageable clinical load	Context-aware thresholds, personalized models	Efficient do pat
Privacy vs Data Utility	Local data retention vs population-scale analytics	On-device inference, selective transmission, federated updates	Enhanced and r com

Sensitivity vs specificity

Screening for undiagnosed AFib demands high sensitivity to capture fleeting paroxysmal episodes, yet excessive false positives erode specificity and clinical confidence [13]. Threshold tuning on edge devices must therefore remain dynamically adjustable, balancing population-level detection goals against individual user tolerance [14]. Studies of consumer wearables have shown that fixed operating points often fail across age groups and activity levels, underscoring the need for personalized decision boundaries [15]. Edge AI uniquely enables on-device recalibration without cloud dependency, opening new avenues for context-aware specificity [16].

The sensitivity-specificity frontier also influences downstream resource utilization, from unnecessary Holter referrals to missed opportunities for stroke prevention [17]. Adaptive inference engines that shift thresholds based on real-time heart-rate variability or activity context can mitigate these tensions [18]. As longitudinal data accumulate, on-device learning will allow models to converge toward user-specific optima, improving both metrics simultaneously [19]. This capability positions edge AI as the enabling technology for precision arrhythmia screening that respects both clinical rigor and human factors [22].

Future Capabilities

Beyond AFib: other arrhythmias

Edge AI on smartwatches will soon extend beyond AFib to detect bradycardia, tachycardia, premature ventricular contractions, and long-QT patterns using the same PPG and ECG pipelines [23]. Generalization across arrhythmia subtypes requires models robust to overlapping morphologies and variable signal quality, challenges that cloud systems address through sheer scale but edge devices must solve through architectural innovation [24]. Early research into multi-label classification on resource-constrained hardware already hints at feasibility, promising a single wearable that serves as a comprehensive rhythm guardian [25]. The clinical payoff could be substantial, shifting from single-disease detection to holistic cardiac surveillance [26].

Future devices may further incorporate burden metrics—such as percentage of time spent in arrhythmia—calculated entirely on-device for immediate clinical relevance [27]. Such expansions will demand careful validation against gold-standard ECG to maintain trust across the expanded label set [7]. Edge architectures that support incremental learning of new arrhythmia classes without full retraining will accelerate this evolution [8]. When realized, these capabilities will transform smartwatches into versatile diagnostic companions rather than narrow AFib detectors [9].

Multimodal edge AI

Fusing PPG, single-lead ECG, accelerometer, and gyroscope data directly on the smartwatch enables superior artifact rejection and richer contextual understanding than any single modality [10]. On-device multimodal fusion reduces reliance on post-hoc cloud processing, delivering cleaner rhythm classifications with lower latency and power overhead [11]. Hardware-efficient attention mechanisms and sensor-fusion layers are already being optimized for tiny memory footprints, paving the way for always-on, context-aware monitoring [12]. The resulting systems will distinguish true arrhythmias from motion-induced noise with unprecedented reliability [13].

Multimodal edge AI also unlocks novel physiological insights, such as activity-stratified arrhythmia risk or sleep-stage influences on rhythm stability [14]. By keeping fusion local, privacy is preserved while enabling real-time adaptive sampling of the most informative sensors [15]. Future platforms will likely treat sensor orchestration as a learned policy rather than a static schedule, further optimizing power and accuracy [16]. This holistic sensing paradigm positions edge AI as the cornerstone of next-generation wearable cardiology [17].

Personalized models

On-device fine-tuning using an individual's own historical PPG and ECG data can dramatically improve model performance without ever transmitting raw waveforms [18]. Privacy-preserving personalization techniques, including federated-style updates executed locally, will allow smartwatches to adapt to unique heart-rate patterns, skin tone, and activity profiles [19]. Such user-specific models reduce both false positives and false negatives compared with population-level baselines, enhancing clinical utility for diverse demographics [22]. The shift toward personalized edge AI marks a departure from one-size-fits-all algorithms that have dominated early wearable deployments [23].

Longitudinal personalization also enables continuous model improvement across years of wear, capturing age-related or disease-progression changes in cardiac signals [24]. Edge hardware must therefore support efficient incremental learning without exhausting battery or memory resources [25]. When combined with secure on-device storage of user embeddings, this capability will deliver arrhythmia detection that feels bespoke rather than generic [26]. Personalized edge AI thus represents the ultimate convergence of consumer convenience and precision medicine [27].

Regulatory and Equity Considerations

FDA clearance for edge AI updates

Over-the-air updates to edge AI models introduce regulatory complexity because each new weight set potentially alters clinical performance, necessitating re-validation under evolving FDA frameworks for software as a medical device [7]. Locked versus adaptive models pose distinct challenges: static models simplify clearance yet limit responsiveness, while continuously learning edge systems require novel post-market surveillance strategies [8]. Early clearance pathways for consumer AFib detectors provide a foundation, yet edge-specific guidance remains nascent [9]. Harmonized international standards will be essential to avoid stifling innovation while safeguarding patient safety [10].

Manufacturers must therefore invest in transparent, hardware-in-the-loop testing pipelines that demonstrate continued safety after deployment [11]. Prospective monitoring of real-world edge AI performance will generate the evidence regulators demand for iterative approvals [12]. The field's ability to balance rapid iteration with rigorous oversight will determine how quickly advanced on-device algorithms reach patients [13]. Ultimately, regulatory agility around edge AI will either accelerate or bottleneck the transition to truly intelligent wearables [14].

Access and equity

Smartwatches remain costly, creating socioeconomic barriers that could exacerbate disparities in AFib screening and stroke prevention [15]. Training data biases toward lighter skin tones and younger, healthier cohorts further risk poorer performance for underrepresented populations when models run on-device [16]. Edge AI itself cannot magically resolve these upstream data inequities, yet it can mitigate some downstream effects through local adaptation to individual physiology [17]. Deliberate efforts in diverse dataset curation and inclusive model evaluation are therefore non-negotiable [18].

Equity also extends to global access, where connectivity-dependent cloud solutions disadvantage low-resource settings far more than fully autonomous edge AI [19]. Low-cost wearable platforms incorporating efficient neural engines could democratize advanced arrhythmia detection if paired with open-source compression toolkits [22]. Policymakers and developers must therefore co-design deployment strategies that prioritize affordability and bias mitigation from the outset [23]. Only then will edge AI fulfill its promise as a universally accessible tool for cardiac health equity [24].

Conclusion

The transition from cloud-centric to fully on-device edge AI for smartwatch-based AFib detection marks a fundamental architectural leap in wearable cardiology. By localizing inference, these systems eliminate latency, preserve privacy, and dramatically improve power efficiency while enabling continuous, real-time monitoring that earlier hybrid approaches could not sustain. Clinical validation efforts have laid the groundwork, yet the true test lies in scaling edge intelligence across millions of devices without compromising usability or safety. This evolution positions smartwatches as proactive clinical partners rather than passive data loggers.

Key challenges in real-time processing, power efficiency, and clinical integration remain tightly intertwined, demanding solutions that treat accuracy, latency, and battery life as co-equal objectives. Model compression, hardware acceleration, and adaptive inference have already demonstrated feasibility, yet their successful orchestration will define the next generation of wearable AFib tools. Regulatory and equity considerations add further layers of complexity that cannot be solved by technology alone. Multidisciplinary collaboration across AI, cardiology, and health policy is therefore indispensable.

Multi-objective optimization emerges as the central design problem confronting the field, moving beyond accuracy-only benchmarks to holistic evaluation protocols that reflect real-world wearable constraints. Standardized on-device benchmarks and open datasets will accelerate progress by enabling fair comparisons across research groups and manufacturers. As these practices take hold, the community can focus on outcomes that matter most to patients—stroke prevention, quality of life, and equitable access. The coming years will reveal which architectures best navigate this complex optimization landscape.

The field must move beyond accuracy-only benchmarking and adopt multi-objective evaluation (accuracy, latency, power). Standardized benchmarks on wearable hardware are urgently needed. Collaborative ecosystems spanning engineers, clinicians, and regulators will determine whether edge AI realizes its full potential or remains an academic curiosity. In the end, the true measure of success will be the silent prevention of strokes through intelligent, always-available rhythm monitoring embedded in everyday life.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 11 Feb 2022 Revised: 28 May 2022 Accepted: 27 Jun 2022
Published online: 20 January 2023

References

Wijesurendra RS, Casadei B. Mechanisms of atrial fibrillation. Heart. 2019;105(24):1860-7. https://doi.org/10.1136/heartjnl-2018-314267.

Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol. 2018;3(5):409-16. https://doi.org/10.1001/jamacardio.2018.0136.

Perez MV, Mahaffey KW, Hedlin H, Rumsfeld JS, Garcia A, Ferris T, et al. Large-scale assessment of a smartwatch to identify atrial fibrillation. N Engl J Med. 2019;381(20):1909-17. https://doi.org/10.1056/NEJMoa1901183.

Bashar SK, Han D, Hajeb-Mohammadalipour S, Ding E, Whitcomb C, McManus DD, et al. Atrial fibrillation detection from wrist photoplethysmography signals using smartwatches. Sci Rep. 2019;9(1):15054. https://doi.org/10.1038/s41598-019-49006-3.

Andersen RS, Peimankar A, Puthusserypady S. A deep learning approach for real-time detection of atrial fibrillation. Expert Syst Appl. 2019;115:465-73. https://doi.org/10.1016/j.eswa.2018.08.011.

Corino VD, Laureanti R, Ferranti L, Scarpini G, Lombardi F, Mainardi LT. Detection of atrial fibrillation episodes using a wristband device. Physiol Meas. 2017;38(5):787-99.

Shen Y, Voisin M, Aliamiri A, Avati A, Hannun A, Ng A. Ambulatory atrial fibrillation monitoring using wearable photoplethysmography with deep learning. In: Proc 25th ACM SIGKDD Int Conf Knowl Discov Data Min. 2019. p. 1909-16. https://doi.org/10.1145/3292500.3330679.

Aliamiri A, Shen Y. Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor. In: 2018 IEEE EMBS Int Conf Biomed Health Inform (BHI). 2018. p. 442-5. https://doi.org/10.1109/BHI.2018.8333471.

Kwon S, Hong J, Choi EK, Lee B, Baik C, Lee E, et al. Detection of atrial fibrillation using a ring-type wearable device (CardioTracker) and deep learning analysis of photoplethysmography signals: prospective observational proof-of-concept study. J Med Internet Res. 2020;22(5):e16443. https://doi.org/10.2196/16443.

Jeon E, Oh K, Kwon S, Son H, Yun Y, Jung ES, et al. A lightweight deep learning model for fast electrocardiographic beats classification with a wearable cardiac monitor: development and validation study. JMIR Med Inform. 2020;8(3):e17037. https://doi.org/10.2196/17037.

Avram R, Ramsis M, Cristal AD, Nathan V, Zhu L, Kim J, et al. Validation of an algorithm for continuous monitoring of atrial fibrillation using a consumer smartwatch. Heart Rhythm. 2021;18(9):1482-90. https://doi.org/10.1016/j.hrthm.2021.05.003.

Badertscher P, Lischer M, Mannhart D, Knecht S, Isenegger C, de Lavallaz JD, et al. Clinical validation of a novel smartwatch for automated detection of atrial fibrillation. Heart Rhythm O2. 2022;3(2):208-10. https://doi.org/10.1016/j.hroo.2022.01.001.

Roselli C, Rienstra M, Ellinor PT. Genetics of atrial fibrillation in 2020: GWAS, genome sequencing, polygenic risk, and beyond. Circ Res. 2020;127(1):21-33. https://doi.org/10.1161/CIRCRESAHA.120.316341.

Campo D, Elie V, de Gallard T, Bartet P, Morichau-Beauchant T, Genain N, et al. Atrial fibrillation detection with an analog smartwatch: prospective clinical study and algorithm validation. JMIR Form Res. 2022;6(11):e37280. https://doi.org/10.2196/37280.

Sabbadini R, Riccio M, Maresca L, Irace A, Breglio G. Atrial fibrillation detection by means of edge computing on wearable device: a feasibility assessment. In: 2022 IEEE Int Symp Med Meas Appl (MeMeA). 2022. p. 1-6. https://doi.org/10.1109/MeMeA54994.2022.9856584.

Gonzalez-Carabarin L, Schmid A, Van Sloun RJ. Hardware-oriented pruning and quantization of deep learning models to detect life-threatening arrhythmias. In: 2021 IEEE Biomed Circuits Syst Conf (BioCAS). 2021. p. 1-6. https://doi.org/10.1109/BioCAS52823.2021.9644947.

Xiaolin L, Panicker RC, Cardiff B, John D. Multistage pruning of CNN based ECG classifiers for edge devices. In: 2021 43rd Annu Int Conf IEEE Eng Med Biol Soc (EMBC). 2021. p. 1965-8. https://doi.org/10.1109/EMBC46164.2021.9630314.

Lu J, Liu D, Cheng X, Wei L, Hu A, Zou X. An efficient unstructured sparse convolutional neural network accelerator for wearable ECG classification device. IEEE Trans Circuits Syst I Regul Pap. 2022;69(11):4572-82. https://doi.org/10.1109/TCSI.2022.3198994.

Lattanzi E, Donati M, Freschi V. Exploring artificial neural networks efficiency in tiny wearable devices for human activity recognition. Sensors (Basel). 2022;22(7):2637. https://doi.org/10.3390/s22072637.

Lubitz SA, Faranesh AZ, Selvaggi C, Atlas SJ, McManus DD, Singer DE, et al. Detection of atrial fibrillation in a large population using wearable devices: the Fitbit heart study. Circulation. 2022;146(19):1415-24. https://doi.org/10.1161/CIRCULATIONAHA.122.060291.

Faust O, Ciaccio EJ, Acharya UR. A review of atrial fibrillation detection methods as a service. Int J Environ Res Public Health. 2020;17(9):3093. https://doi.org/10.3390/ijerph17093093.

Hooshmand M, Zordan D, Melodia T, Rossi M. SURF: Subject-adaptive unsupervised ECG signal compression for wearable fitness monitors. IEEE Access. 2017;5:19517-35. https://doi.org/10.1109/ACCESS.2017.2754683.

Meng L, Tan W, Ma J, Wang R, Yin X, Zhang Y. Enhancing dynamic ECG heartbeat classification with lightweight transformer model. Artif Intell Med. 2022;124:102236. https://doi.org/10.1016/j.artmed.2022.102236.

Wang F, Ma Q, Liu W, Chang S, Wang H, He J, et al. A novel ECG signal compression method using spindle convolutional auto-encoder. Comput Methods Programs Biomed. 2019;175:139-50. https://doi.org/10.1016/j.cmpb.2019.04.015.

Nunez-Yanez J, Howard N. Energy-efficient neural networks with near-threshold processors and hardware accelerators. J Syst Archit. 2021;116:102062. https://doi.org/10.1016/j.sysarc.2021.102062.

Essien UR, Kornej J, Johnson AE, Schulson LB, Benjamin EJ, Magnani JW. Social determinants of atrial fibrillation. Nat Rev Cardiol. 2021;18(11):763-73. https://doi.org/10.1038/s41569-021-00561-7.

Magno M, Pritz M, Mayer P, Benini L. DeepEmote: Towards multi-layer neural networks in a low power wearable multi-sensors bracelet. In: 2017 7th IEEE Int Workshop Adv Sensors Interfaces (IWASI). 2017. p. 32-7. https://doi.org/10.1109/IWASI.2017.7974205.