

AI-Driven Detection of Cardiovascular Abnormalities from Wearable Sensor Data

Hugo M. Hansen

Department of Computer Science, George Mason University, Fairfax, VA, USA.
hansenhugo@gmu.edu

Juanye Sun

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

Antonio Woods

Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence,
KS, USA.
antoniowoods815@ku.edu

Abstract

Cardiovascular diseases remain a leading cause of mortality, and wearable sensors offer continuous physiological streams that can transform detection, monitoring, and risk stratification. This paper presents an interdisciplinary systems analysis of AI-driven detection of cardiovascular abnormalities from wearable sensor data. It examines the clinical rationale, data characteristics, model architectures, and system-level trade-offs that define contemporary implementations. Rather than focusing solely on algorithmic accuracy, the analysis emphasizes infrastructure, deployment, robustness, fairness, governance, and sustainability. The discussion integrates perspectives from biomedical signal processing, distributed systems, machine learning, and health policy. Wearable data exhibit heterogeneity, motion sensitivity, and contextual noise, requiring substantial preprocessing and uncertainty management. Deep learning architectures can learn morphological and temporal patterns, but their clinical utility depends on integration with real-time infrastructure, interoperability, and rigorous prospective evaluation. Structural trade-offs include edge versus cloud processing, model complexity versus explainability, and personalization versus generalizability. Federated learning and decentralized architectures offer privacy-preserving routes for multi-institutional training. At the same time, algorithmic bias, ethical oversight, and governance remain critical concerns. The paper argues that sustainable implementations require lifecycle planning, energy-aware design, regulatory alignment, and equitable access. Forward-looking perspectives address multimodal fusion, adaptive calibration, and policy incentives for responsible deployment. Overall, the analysis frames AI-driven wearable cardiovascular detection as a socio-technical system in which reliable clinical benefit emerges from careful alignment among sensing, learning, infrastructure, and governance.

Keywords

artificial intelligence; cardiovascular abnormalities; wearable sensors; digital health; machine learning; health systems.

1. Introduction

Cardiovascular disease continues to impose a substantial clinical and societal burden, representing a leading cause of death and disability across many populations [1]. At the same time, wearable devices have matured from lifestyle accessories into increasingly credible instruments for ambulatory physiological monitoring, enabling continuous or near-continuous observation of cardiac rhythm, heart rate, oxygen saturation, physical activity, and related signals [2]. The resulting data streams create new opportunities to detect cardiovascular abnormalities earlier and to monitor disease progression outside conventional clinical settings. However, these opportunities are accompanied by significant technical, clinical, and organizational challenges. Wearable sensing platforms are heterogeneous in terms of sensor quality, sampling rates, body placement, connectivity, and patient adherence. The integration of artificial intelligence into these environments therefore requires careful attention to the full socio-technical system rather than a narrow focus on model accuracy.

From an engineering perspective, wearable cardiovascular monitoring involves the acquisition of photoplethysmographic signals, electrocardiographic traces, accelerometry, and sometimes impedance-derived or acoustic information [3]. The analysis of these signals has historically relied on heuristic feature extraction and threshold-based decision rules. More recently, data mining and machine learning methods have been applied to wearable health data to identify patterns that may escape manual inspection and to support continuous risk assessment [4]. These approaches can be trained to recognize transient arrhythmias, evidence of ischemic burden, and hemodynamic instability, but their clinical translation depends on robust data quality management, interpretability, and integration into existing care pathways.

This paper adopts a system-level perspective on AI-driven detection of cardiovascular abnormalities from wearable sensor data. It examines the clinical context, the structural characteristics of wearable data, the main families of learning architectures, and the trade-offs that arise during deployment. The analysis extends beyond model performance to include infrastructure design, robustness, fairness, governance, and sustainability. In doing so, it positions AI-based detection as a complex intervention that interacts with clinical workflows, regulatory frameworks, and patient behavior. The central argument is that meaningful progress requires coherence across sensing technology, algorithmic design, deployment infrastructure, and policy.

2. Background and Clinical Context

The clinical motivation for continuous cardiovascular monitoring arises from the episodic nature of many important abnormalities. Atrial fibrillation, ventricular ectopy, conduction disturbances, and ischemia may be intermittent, asymptomatic, or triggered by physical or emotional stress. Traditional Holter monitoring and event recorders provide limited observation windows and may miss events that occur outside the monitoring period. In contrast, wearable systems can extend observation over days, weeks, or longer, producing a more complete picture of cardiovascular behavior. However, ambulatory monitoring also introduces variability from movement, skin condition, sensor displacement, environmental factors, and patient-specific physiology. These sources of variability create significant challenges for reliable signal interpretation [5].

At the algorithmic level, deep neural networks have demonstrated impressive capability in detecting cardiac rhythm abnormalities from ambulatory electrocardiograms. For example, end-to-end convolutional architectures have achieved cardiologist-level performance for certain arrhythmia classification tasks when evaluated on large annotated datasets [6]. Related work has shown that AI-enabled interpretation of standard electrocardiographic signals can

identify patterns associated with reduced cardiac contractile function, suggesting that machine learning may uncover subclinical structural abnormalities from electrical data [7]. In the wearable context, commercially available smartwatches using photoplethysmographic sensors have been investigated for passive detection of atrial fibrillation, illustrating the potential of non-electrocardiographic signals for population-level screening [8]. These examples show that the clinical promise of AI is not limited to a single sensor modality or disease class.

The broader significance of these advances has been discussed as part of the maturation of deep learning in healthcare, where data availability, computational capacity, and algorithmic innovation have converged to enable new forms of diagnostic support [9]. At the same time, the concept of high-performance medicine emphasizes that artificial intelligence should augment human clinicians by improving speed, consistency, and the capacity to process large volumes of data [10]. However, these advances do not automatically translate into safe and effective wearable deployments. The path from a retrospective model evaluation to a reliable clinical service involves substantial additional engineering, validation, and governance.

3. Wearable Data Characteristics and Preprocessing

Wearable cardiovascular data differ from hospital-grade data in several important respects. Signals are often acquired at lower sampling rates, with limited electrode contact or optical coupling, and under conditions of unrestrained movement. Photoplethysmographic signals, for example, are highly sensitive to motion artifacts, ambient light, skin pigmentation, and peripheral perfusion. These effects can distort pulse morphology and complicate the extraction of heart rate, heart rate variability, and surrogates of blood pressure or vascular tone. The challenges become more severe during varying physical activity intensity, because physiological changes and motion-induced noise are simultaneously present. In this context, methods for rapid vital sign extraction must be designed to preserve physiological information while suppressing nonstationary interference [11].

Preprocessing therefore occupies a critical role in the overall system. Signal conditioning typically includes filtering, artifact detection, normalization, and segmentation. Quality assessment is particularly important because not all segments of a long wearable recording are equally informative. Segments contaminated by motion or sensor detachment may be assigned lower weights or excluded from analysis. In addition, data from multiple sensors may need to be aligned temporally and resampled to a common reference. Multimodal fusion can improve robustness by combining complementary information, but it also increases system complexity and may introduce new failure modes. The design of preprocessing pipelines thus involves trade-offs between sensitivity, specificity, computational load, and the ability to operate in real time on constrained wearable or edge devices.

Another central challenge is the management of contextual and behavioral information. Physical activity, sleep, stress, medication adherence, and environmental conditions can all influence cardiovascular signals. Contextual data are valuable for interpreting observed changes and reducing false positives. For example, a transient increase in heart rate during exercise has different clinical significance than the same increase during rest. Integrating these contextual cues requires data architectures that can handle heterogeneous sources, missing values, and irregular sampling. From a systems perspective, preprocessing is not merely a technical step but a boundary through which raw sensor streams are transformed into meaningful physiological evidence.

4. AI Detection Architectures

A range of machine learning architectures has been applied to cardiovascular signal analysis. Convolutional neural networks are widely used to learn local morphological features from electrocardiographic and photoplethysmographic waveforms. Patient-specific one-dimensional convolutional networks, for instance, can be adapted to individual cardiac morphology and have demonstrated strong real-time performance in ECG classification tasks [12]. Recurrent neural networks and temporal convolutional networks are also used to capture dependencies across heartbeats and longer temporal contexts. Attention mechanisms can help models focus on clinically relevant segments, while multimodal architectures combine electrocardiographic, photoplethysmographic, and accelerometry inputs to improve robustness under ambulatory conditions.

The choice of architecture is closely linked to the detection task and the deployment environment. For continuous monitoring, models must process long recordings efficiently, often using sliding windows or event-triggered inference. For embedded deployment, low complexity and memory footprint are essential. For cloud-based retrospective analysis, larger models and ensembles may be more acceptable. In many cases, a hybrid strategy is adopted in which a lightweight model performs initial screening on the device and a more complex model conducts confirmatory analysis in the cloud. This type of staged inference reflects a broader systems engineering principle: computational resources should be allocated according to the prior probability and clinical urgency of the event.

Architectural decisions also interact with the availability of labeled data. Cardiovascular abnormalities are often rare, and high-quality annotations are costly to obtain. Self-supervised pretraining, transfer learning, and data augmentation can improve performance when annotated examples are limited. However, these techniques introduce additional assumptions about the relationship between pretraining and target tasks. Moreover, model complexity must be balanced against the need for transparency and auditable behavior. The selection of an architecture is therefore not simply a technical optimization problem but a decision that shapes the reliability, maintainability, and clinical acceptance of the system.

5. System-Level Trade-offs

The deployment of AI-driven cardiovascular detection involves multiple structural trade-offs. One fundamental tension arises between accuracy and interpretability. Deep learning models may achieve high predictive performance while remaining difficult to explain in mechanistic terms. Techniques for local explanation, such as model-agnostic methods that highlight influential input segments, have been proposed to support clinical understanding [13]. However, explanation methods themselves can be unstable and may not fully capture the reasoning of a complex model. Clinicians often require an explanation that aligns with physiological knowledge, yet current saliency-based approaches do not always provide such alignment. System designers must decide how to present uncertainty, confidence scores, and contextual factors in ways that support rather than undermine clinical judgment.

A second trade-off concerns centralization and decentralization. Cloud-based processing enables powerful analytics, centralized model updates, and integration with electronic health records. However, it depends on reliable network connectivity and raises privacy concerns. Edge processing reduces latency and supports offline operation, but it limits model size and complicates model maintenance. Federated learning offers an intermediate path by allowing multiple institutions or devices to contribute model updates without sharing raw data [14]. This approach can improve privacy and enable learning across distributed populations, but it

introduces challenges related to communication efficiency, data heterogeneity, and incentive alignment.

A third trade-off involves personalization versus generalizability. A model trained on a broad population may not perform optimally for individuals with unusual cardiac morphology or rare conditions. Personalized models can adapt to individual baselines, but they require additional data, storage, and regulatory consideration. The appropriate balance depends on the clinical use case. For screening applications, generalizable models may be preferable. For long-term monitoring of a known patient, adaptive personalization may reduce false alarms and improve detection sensitivity. These trade-offs illustrate that performance metrics alone are insufficient to guide design; the operational context and downstream consequences must also be considered.

6. Deployment and Infrastructure

Deploying AI-driven wearable cardiovascular detection at scale requires more than a trained model. The infrastructure must support data ingestion, real-time or near-real-time inference, alert generation, storage, and integration with clinical systems. Scalable architectures developed for electronic health records provide useful lessons for organizing these components into modular pipelines [15]. Such pipelines typically separate data collection, feature extraction, model inference, and decision support, allowing independent evolution of each layer. In wearable deployments, this separation is particularly important because sensors and devices evolve rapidly, and the clinical decision logic may need to be updated without replacing the entire system.

The reliability of the deployment infrastructure is closely linked to patient safety. A missed alert may delay treatment, while excessive false alarms may lead to alarm fatigue and reduced adherence. Therefore, the system must be designed to handle uncertain connectivity, device battery limitations, and varying data quality. Alerting policies should consider the severity of the suspected abnormality, the confidence of the model, and the availability of clinical response pathways. Prospective evaluation of machine learning-based decision support is essential to understand how these systems affect clinician behavior and patient outcomes [16]. Without such evaluation, retrospective performance metrics may overstate real-world benefit.

Interoperability also plays an important role. Wearable data must be mapped to standardized formats and integrated with electronic health records if they are to support clinical decision-making. This requires alignment with health information exchange standards, data governance policies, and institutional workflows. In practice, many health systems struggle with data silos, inconsistent device identifiers, and incomplete metadata. Addressing these issues is not merely a technical exercise; it requires organizational agreements about data ownership, consent, and accountability. The infrastructure design must therefore include not only computational components but also the institutional arrangements that make the data meaningful and actionable.

7. Robustness and Generalization

Robustness across populations, devices, and environments is a central challenge for AI-based wearable cardiovascular detection. Models trained on data from a specific demographic group or device type may perform poorly when applied elsewhere. Variation in skin tone, body composition, sensor quality, and wear position can all influence signal characteristics. Disease prediction models developed over large healthcare community datasets have shown promise, yet their generalization to new settings requires careful validation [17]. Without representative

training data and systematic evaluation across subpopulations, models may encode hidden biases or fail silently under distribution shift.

Robustness also depends on the ability to detect and respond to data quality problems at runtime. Sensor detachment, battery failure, and network interruptions can create missing or corrupted data. A robust system should be able to recognize these conditions and suppress unreliable predictions. This may involve uncertainty quantification, out-of-distribution detection, and confidence calibration. However, these techniques add computational and architectural complexity. The design of robustness mechanisms must be proportionate to the clinical risk associated with each prediction. For example, screening applications that are later confirmed by clinical evaluation may tolerate higher uncertainty than automatic alarm systems that trigger immediate intervention.

The capacity to learn from distributed data without compromising privacy is another factor in generalization. Federated learning can enable model development across multiple institutions and populations, thereby improving representativeness while preserving data governance [18]. Yet federated systems face challenges related to heterogeneous data distributions, device capabilities, and the coordination of training rounds. These challenges are not purely technical; they also involve trust among participating organizations and the development of shared evaluation norms. A federated infrastructure must be governed by clear protocols for model governance, contribution attribution, and safety monitoring.

8. Fairness, Governance, and Policy

Fairness in AI-driven cardiovascular detection has both technical and normative dimensions. Research on algorithmic fairness has shown that predictive models used in healthcare can encode and amplify existing disparities when trained on biased data or when proxy outcomes reflect structural inequities [19]. In wearable monitoring, disparities may arise from differences in device access, sensor accuracy across skin tones, or uneven representation in training datasets. If a model is less accurate for certain groups, the resulting false positives or missed detections can produce differential clinical harm. Fairness therefore requires active measurement and mitigation rather than passive assumption of neutrality.

Ethical challenges in implementing machine learning in health care include transparency, accountability, informed consent, and the appropriate allocation of clinical authority [20]. Wearable AI systems introduce additional questions about continuous surveillance, patient autonomy, and the management of incidental findings. These systems may generate large volumes of data that extend beyond the original clinical question. Institutions must define how such findings are communicated, who is responsible for follow-up, and how patient preferences are respected. Without clear governance structures, the technology may create new forms of uncertainty and burden for both patients and clinicians.

Responsible machine learning frameworks emphasize the importance of prospective validation, stakeholder engagement, auditability, and ongoing monitoring [21]. Such frameworks recognize that model performance is not static and that clinical environments evolve over time. Governance mechanisms should include routine performance review, adverse event reporting, and clear criteria for model retraining or decommissioning. Global guidance on ethics and governance of artificial intelligence for health similarly highlights transparency, equity, human oversight, and accountability as foundational principles [22]. These principles must be translated into operational policies, institutional review processes, and regulatory requirements that are appropriate for wearable and consumer-facing devices.

9. Sustainability and Lifecycle Considerations

Sustainability is often overlooked in discussions of AI-enabled health systems, yet it is essential for long-term clinical and economic viability. Wearable solutions require ongoing investment in device maintenance, software updates, data storage, and personnel training. The environmental footprint of data centers and network infrastructure also deserves attention, particularly when continuous monitoring generates large volumes of physiological data. A sustainable deployment must therefore balance clinical benefit against resource consumption. Lifecycle planning should include device disposal, data retention policies, and the eventual decommissioning of models or services.

From a technical perspective, sustainable systems require modularity and adaptability. Sensor hardware evolves quickly, and the ability to update algorithms without replacing the entire platform can reduce long-term costs. Application programming interfaces, data standards, and versioned model registries can support incremental change. In addition, the energy consumption of on-device inference should be managed to avoid excessive battery drain and user attrition. Edge computing approaches can reduce network traffic but may require hardware-specific optimization. These considerations illustrate that sustainability links technical design choices with patient adherence and system longevity.

The long-term success of AI-driven wearable cardiovascular monitoring also depends on workforce readiness and institutional capacity. Clinicians need training to interpret model outputs, understand limitations, and communicate uncertainties to patients in a way that supports shared decision-making and preserves trust. Without such training, the introduction of wearable AI may increase cognitive burden and create confusion about the appropriate level of clinical concern. Institutional protocols should define escalation pathways that remain responsive to model outputs while preserving clinical judgment. The sustainability of these systems also depends on their alignment with reimbursement models and the value they provide to both patients and health systems. Devices that reduce avoidable admissions or enable earlier intervention may generate long-term savings, but these benefits must be demonstrated through rigorous health economic evaluations. In the absence of such evidence, adoption may be limited to pilot programs and technology enthusiasts, rather than becoming a standard component of cardiovascular care. Lifecycle planning must therefore account for the gradual accumulation of evidence, the need for model updates as populations and devices change, and the eventual replacement of hardware and software. A sustainable deployment is one that can adapt to these changes without disrupting clinical services or compromising patient safety.

10. Conclusion

AI-driven detection of cardiovascular abnormalities from wearable sensor data is best understood as a socio-technical system in which clinical value emerges from the coordinated design of sensing, learning, infrastructure, and governance. The analysis in this paper has emphasized that model accuracy, while necessary, is insufficient to guarantee real-world benefit. Wearable data possess unique characteristics related to motion, physiology, context, and device heterogeneity. These characteristics demand careful preprocessing, robust uncertainty management, and adaptation to the constraints of ambulatory deployment. Architectures such as convolutional neural networks, recurrent models, and multimodal transformers can extract clinically relevant patterns, but their success depends on integration with scalable infrastructure, prospective evaluation, and clear pathways for clinical response.

Structural trade-offs shape every major design decision. Edge processing offers lower latency and greater privacy but limits model complexity. Cloud processing enables more sophisticated inference but depends on connectivity and central data governance. Federated learning allows distributed training without raw data sharing, yet introduces coordination and heterogeneity challenges. Personalization can improve accuracy for individual patients but complicates regulatory oversight and long-term maintenance. Explainability is essential for clinical trust, yet current methods often fail to align with physiological reasoning. Each of these trade-offs must be resolved within the specific context of the intended clinical use case, the available infrastructure, and the regulatory environment.

Fairness and governance are not peripheral concerns but foundational requirements for responsible deployment. Bias can arise from unrepresentative training data, differential sensor performance, and structural inequities in access to wearable technology. Without proactive fairness auditing, AI systems may exacerbate existing disparities in cardiovascular care. Governance frameworks must address accountability, informed consent, incidental findings, and the allocation of clinical authority between algorithms and humans. External guidance documents and responsible AI principles provide useful direction, but their implementation requires institutional commitment, audit mechanisms, and stakeholder engagement. Patient and clinician perspectives should inform the entire lifecycle, from problem definition to post-deployment monitoring.

Looking forward, the field is likely to benefit from several converging developments. Multimodal sensing will integrate cardiac, respiratory, activity, and contextual signals to provide a richer view of cardiovascular health. Self-supervised learning and adaptive calibration will allow models to maintain performance as devices and populations evolve. Real-time infrastructure will continue to mature, enabling reliable alerting, explainable decision support, and integration with electronic health records. At the same time, policy incentives must be aligned to reward not only technical novelty but also clinical validity, equity, and long-term sustainability. Achieving this alignment requires continued interdisciplinary collaboration among clinicians, engineers, data scientists, ethicists, and health system leaders. The future of AI-driven cardiovascular monitoring depends less on any single algorithmic breakthrough than on the ability to build and govern systems that are robust, fair, transparent, and sustainable across diverse real-world settings.

References

1. Virani, S. S., Alonso, A., Benjamin, E. J., Bittencourt, M. S., Callaway, C. W., Carson, A. P., ... & American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. (2020). Heart disease and stroke statistics—2020 update: A report from the American Heart Association. *Circulation*, 141(9), e139–e596.
2. Piwek, L., Ellis, D. A., Andrews, S., & Joinson, A. (2016). The rise of consumer health wearables: Promises and barriers. *PLOS Medicine*, 13(2), e1001953.
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. Witt, D. R., Kellogg, R. A., Snyder, M. P., & Dunn, J. (2019). Windows into human health through wearables data analytics. *Current Opinion in Biomedical Engineering*, 9, 28–46.

5. Bansal, A., & Joshi, R. (2018). Portable out-of-hospital electrocardiography: A review of current technologies. *Journal of Arrhythmia*, 34(2), 129–138.
6. 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.
7. Attia, Z. I., Kapa, S., Lopez-Jimenez, F., McKie, P. M., Ladewig, D. J., Satam, G., ... & Friedman, P. A. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*, 25(1), 70–74.
8. Perez, M. V., Mahaffey, K. W., Hedlin, H., Rumsfeld, J. S., Garcia, A., Ferris, T., ... & Turakhia, M. P. (2019). Large-scale assessment of a smartwatch to identify atrial fibrillation. *New England Journal of Medicine*, 381(20), 1909–1917.
9. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29.
10. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
11. Zheng, X., Dwyer, V. M., Barrett, L. A., Derakhshani, M., & Hu, S. (2023). Rapid vital sign extraction for real-time opto-physiological monitoring at varying physical activity intensity levels. *IEEE Journal of Biomedical and Health Informatics*, 27(7), 3107–3118.
12. Kiranyaz, S., Ince, T., & Gabbouj, M. (2016). Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3), 664–675.
13. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.
14. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 1273–1282.
15. Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association*, 23(5), 899–908.
16. Beede, E., Baylor, E., Hersch, F., Iurchenko, A., Wilcox, L., Ruamviboonsuk, P., ... & Webster, D. R. (2020). A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12.
17. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., ... & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1, 18.
18. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3, 119.

19. 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.
20. Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *New England Journal of Medicine*, 378(11), 981–983.
21. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., ... & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine*, 25(9), 1337–1340.
22. World Health Organization. (2021). Ethics and governance of artificial intelligence for health: WHO guidance. World Health Organization.