AI-Assisted Sleep Quality Assessment Using Wrist-Worn Physiological Sensors

Rahul A. Miles
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
rahul_a_miles@binghamton.edu

Pengfu Meng
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
hellopengfu@unr.edu

Abstract

Sleep quality assessment is undergoing a significant transformation as wrist-worn physiological sensors become capable of capturing continuous photoplethysmography, accelerometry, and derived autonomic signals in naturalistic environments. Artificial intelligence methods can translate these multimodal sensor streams into clinically meaningful sleep quality indicators, but the design of such systems raises substantial interdisciplinary challenges that extend well beyond model accuracy. This paper presents a systems-level examination of AI-assisted sleep quality assessment using wrist-worn physiological sensors. It analyzes the architectural tensions among on-device processing, cloud-based inference, and clinical integration, and it discusses the sensing, data infrastructure, machine learning, robustness, fairness, governance, and deployment dimensions that jointly determine system viability. The paper argues that effective sleep AI systems must be understood as socio-technical infrastructures rather than isolated algorithmic artifacts. Structural trade-offs include the balance between energy efficiency and signal fidelity, the tension between deep learning representational capacity and interpretability, the need for population diversity in training data, and the regulatory requirement for continuous monitoring of model behavior after deployment. The discussion draws on existing wearable sleep research, clinical artificial intelligence, algorithmic fairness, and health data governance to articulate design considerations for sustainable and trustworthy sleep assessment platforms. By foregrounding system architecture and governance rather than narrow algorithmic performance, the paper offers a cross-domain perspective on building responsible AI-enabled sleep health infrastructures.

Keywords

sleep quality assessment, wearable sensors, artificial intelligence, photoplethysmography, actigraphy, health data governance, algorithmic fairness, clinical deployment.

1. Introduction

Sleep is a fundamental physiological process that affects cognitive function, cardiometabolic health, immune regulation, and emotional well-being. Traditional polysomnography remains the reference standard for sleep evaluation, but it requires specialized laboratories, trained technicians, and controlled conditions that diverge substantially from habitual sleep environments. Wrist-worn physiological sensors offer a complementary pathway by enabling longitudinal monitoring of movement, heart rate, pulse waveform, and derived measures across weeks or months. The growing availability of these devices has stimulated interest in

artificial intelligence systems that can infer sleep quality, sleep architecture, and respiratory disturbances from continuous sensor data. However, translating raw wrist-worn signals into actionable sleep assessments is not simply a signal processing problem. It involves complex system design choices related to sensor placement, data transmission, computational infrastructure, model training, regulatory oversight, and equitable deployment [1].

The clinical and public health promise of wearable sleep assessment lies in its capacity to move sleep measurement from episodic laboratory observation to continuous real-world monitoring. Longitudinal sleep data can reveal patterns that single-night studies cannot capture, including the effects of shift work, medication changes, stress, and behavioral interventions. At the same time, wearable systems introduce uncertainty because they rely on indirect physiological proxies rather than direct neurophysiological measurement. Wrist actigraphy has long been used to estimate sleep and wake states through movement patterns, yet its limitations are well documented, particularly in populations with fragmented sleep or low movement distinction between sleep and quiet wakefulness [3]. The integration of photoplethysmography and machine learning has expanded the information content available from wrist devices, but it has also increased system complexity and created new demands for validation, standardization, and responsible deployment [2].

This paper approaches AI-assisted sleep quality assessment from a systems perspective. Rather than proposing a single algorithmic pipeline, it examines the structural and organizational dimensions that shape how such systems operate in practice. The analysis covers sensing architecture, data infrastructure, analytical modeling, robustness under physiological variability, fairness across populations, privacy and regulatory governance, and operational sustainability. Throughout, the emphasis is placed on trade-offs and interdependencies. For example, model complexity may improve laboratory accuracy but reduce interpretability and increase the risk of silent failure in real-world conditions. Likewise, cloud-based architectures may offer computational flexibility but raise data sovereignty and latency concerns. These trade-offs cannot be resolved through technical optimization alone; they require institutional governance, clinical collaboration, and sustained investment in data quality and diversity [4].

The remainder of the paper is organized as follows. Section 2 reviews the wearable sleep assessment landscape and the role of artificial intelligence. Section 3 examines system architecture and data infrastructure. Section 4 discusses modeling choices and analytical trade-offs. Section 5 addresses robustness and physiological variability. Section 6 considers fairness and generalizability. Section 7 analyzes governance, privacy, and regulatory implications. Section 8 explores deployment and operational sustainability. Section 9 concludes with forward-looking observations.

2. Wearable Sleep Assessment and the Rise of AI-Based Inference

Wrist-worn sleep sensing has evolved from simple actigraphy to multimodal physiological monitoring. Early actigraphic methods estimated sleep and wake by analyzing movement counts with rule-based algorithms that required careful threshold selection and population-specific calibration [3]. More recent open-access approaches have demonstrated that raw accelerometer data can be processed using transparent heuristics to estimate sleep duration in large epidemiological cohorts, thereby improving reproducibility and methodological consistency [4]. These actigraphy-based methods remain valuable because they are computationally efficient and interpretable, but they generally cannot distinguish sleep stages with high resolution and are vulnerable to misclassifying quiet wakefulness as sleep.

The addition of photoplethysmography to wrist-worn devices has substantially enriched the physiological information available for sleep assessment. Photoplethysmography captures pulse wave signals that reflect cardiac activity, vascular tone, and autonomic modulation. When combined with accelerometry, these signals allow machine learning models to estimate sleep stages and sleep quality indicators with improved accuracy compared with movement-only methods [5]. Consumer devices have also been evaluated against polysomnography, with meta-analytic evidence suggesting that multi-sensor wristbands perform better than simple actigraphs but exhibit device-specific variability in sleep stage classification [6]. These findings indicate that AI-assisted sleep assessment should not be treated as a homogeneous category; rather, system performance depends on sensor quality, signal processing, and the specific modeling approach.

Deep learning has accelerated progress in wearable sleep inference by enabling models to learn hierarchical representations from raw or minimally processed sensor data. Transfer learning and domain adaptation have been explored to address differences across devices and patient populations, reducing the need for large labeled polysomnography datasets for every new deployment context [7]. Recurrent neural networks and related sequence models have been used to exploit the temporal structure of sleep, capturing transitions between wake, rapid eye movement sleep, and non-rapid eye movement stages from heart rate variability and body movement signals in clinical populations [8]. These developments illustrate a broader shift from feature engineering based on prior physiological knowledge toward data-driven representation learning. However, this shift also introduces governance challenges because learned representations may obscure failure modes, embed population-specific assumptions, and complicate regulatory review.

From an infrastructure perspective, wearable sleep AI systems are embedded within larger Internet of Things and smart health ecosystems. Sensor data must flow from wearable devices through mobile gateways or home hubs to backend services where data cleaning, synchronization, model inference, and clinical reporting occur [9]. Each infrastructure layer introduces potential points of failure, latency, and privacy risk. The design of these layers must therefore account for data loss during synchronization, battery constraints, network availability, and the need to maintain longitudinal records without overburdening users. A system that performs well in controlled pilot studies may fail in everyday settings if its infrastructure cannot sustain reliable data collection, secure transmission, and timely feedback over months of use. Consequently, AI-assisted sleep quality assessment must be understood as a distributed system that integrates wearable hardware, edge computing, cloud services, clinical interfaces, and human workflows.

3. System Architecture and Data Infrastructure

The architecture of a wearable sleep assessment system determines how data are acquired, processed, transmitted, stored, and transformed into clinically useful outputs. A central design tension exists between on-device processing and cloud-based inference. On-device processing reduces data transmission, improves privacy, and enables real-time feedback even when connectivity is intermittent. However, wearable devices have limited computational resources, battery capacity, and memory, which constrains the complexity of models that can be executed locally. Cloud-based inference allows more sophisticated models and centralized model updates, but it requires continuous connectivity, increases data exposure, and may introduce latency that is problematic for applications such as acute sleep disturbance detection. Hybrid architectures attempt to balance these concerns by performing lightweight signal

quality assessment and feature extraction on the device while delegating computationally intensive inference and longitudinal analysis to the cloud.

The data infrastructure supporting sleep AI must accommodate heterogeneous data streams with different sampling rates, noise characteristics, and temporal alignments. Accelerometry data are often sampled at tens of hertz, while photoplethysmography may be sampled at higher rates depending on the sensor. Heart rate and heart rate variability are typically derived from photoplethysmography through signal processing steps that must be robust to motion artifacts and irregular pulses. Prior work has shown that rapid vital sign extraction can be achieved under varying physical activity intensity using wrist opto-physiological monitoring, which is important because sleep assessment systems often inherit data collected during daytime activity as well as nocturnal rest [10]. The data infrastructure must therefore support contextual awareness, allowing models to distinguish between sleep, rest, and activity periods and to apply appropriate preprocessing strategies.

A further architectural consideration is the need for continuous data quality monitoring. Wrist-worn sensors are susceptible to displacement, poor skin contact, ambient light interference, and device-specific sensor drift. A sleep AI system that ingests raw or lightly processed signals without quality checks may produce confident but incorrect assessments. System architectures therefore increasingly include automated quality assessment modules that evaluate signal coverage, motion artifact burden, pulse waveform morphology, and inter-sensor consistency before data reach the inference stage. These modules may be implemented at the edge or in the cloud depending on trade-offs between responsiveness and computational load. However, quality filtering must be carefully designed to avoid excluding data from populations with atypical signal characteristics, because overly conservative filters can introduce selection bias and reduce generalizability.

Data infrastructure also shapes model lifecycle management. Sleep AI models require periodic retraining or recalibration as sensor hardware evolves, firmware changes, user populations shift, and clinical definitions are updated. A robust architecture includes versioned data pipelines, lineage tracking, and model registries that document which data version trained which model and which model produced which prediction. This type of infrastructure is often invisible in proof-of-concept studies but becomes essential in clinical or commercial deployment. Without it, model updates can inadvertently alter performance in ways that are difficult to detect, especially when the underlying data distribution drifts slowly over time. System architects must therefore treat data provenance, model versioning, and performance auditing as first-class requirements rather than afterthoughts.

4. Modeling Pipeline and Analytical Trade-offs

The modeling pipeline in wearable sleep AI typically includes preprocessing, feature extraction or representation learning, sequential modeling, and postprocessing. Classical approaches rely on handcrafted features such as activity counts, circadian phase markers, heart rate variability metrics, and respiratory rate estimates. These features are interpretable and align with physiological knowledge, but they may discard information that is present in raw waveforms. Deep learning models can learn representations directly from raw or minimally processed sensor data, often achieving higher accuracy in controlled evaluations [7]. However, deep representations may be sensitive to sensor differences, sampling rate changes, and population characteristics, and they may require large amounts of labeled data to generalize reliably.

A central analytical trade-off involves model capacity and clinical interpretability. High-capacity models such as convolutional and recurrent networks can capture complex temporal dependencies in sleep data, but they often function as black boxes. Clinicians may be reluctant to rely on sleep quality scores without understanding which physiological patterns contributed to a given assessment. Hybrid architectures that combine learned representations with interpretable bottleneck features can partially address this concern. For instance, a model may use learned embeddings to estimate sleep stages while also reporting heart rate variability trends, movement epochs, and signal quality indicators that clinicians can review. This dual output structure supports clinical trust and facilitates error investigation without necessarily sacrificing predictive performance.

Another important trade-off relates to personalization and population-level generalization. Sleep physiology varies across age groups, sexes, clinical conditions, and habitual sleep schedules. A global model trained on a heterogeneous sample may perform reasonably for average users but poorly for atypical groups. Personalized models can improve individual accuracy by adapting to each user's baseline patterns, but they introduce additional complexity in model versioning, data storage, and regulatory accountability. Federated learning and on-device adaptation offer potential pathways for personalization while limiting raw data transfer, but these approaches require careful design to ensure that local updates do not overwrite beneficial global knowledge or amplify idiosyncratic noise.

The choice of sleep quality outputs also shapes the modeling pipeline. Some systems aim to estimate discrete sleep stages, while others produce continuous sleep quality indices, restfulness scores, or respiratory disturbance measures. These outputs have different clinical meanings and different validation requirements. Sleep staging is often evaluated against polysomnography-derived hypnograms, but sleep quality perception is influenced by subjective, behavioral, and contextual factors that cannot be fully captured by physiological sensors. AI systems that report sleep quality must therefore distinguish between objective physiological markers and subjective sleep experience, and they should communicate uncertainty when these dimensions diverge. Modeling choices thus have direct implications for how users interpret sleep feedback and how clinicians integrate it into care.

5. Robustness, Artifacts, and Physiological Variability

Robustness is a defining challenge for wrist-worn sleep AI because real-world data contain motion artifacts, environmental noise, and physiological variability that are often underrepresented in training datasets. Motion artifacts are particularly problematic for photoplethysmography-derived features because physical movement can corrupt pulse waveforms and produce misleading heart rate estimates. The ability to extract vital signs reliably across varying physical activity intensity is therefore a critical enabler for sleep systems that collect data continuously across day and night [10]. Sleep-specific processing can sometimes assume lower motion levels, but patients with periodic limb movements, insomnia-related restlessness, or comorbid conditions may still generate substantial nocturnal artifacts.

Physiological variability also complicates model performance. Heart rate dynamics during sleep are influenced by age, medication, alcohol, caffeine, physical fitness, and cardiovascular status. Wrist photoplethysmography signals can be affected by skin tone, tissue composition, peripheral perfusion, and ambient temperature. If training data are not sufficiently diverse, models may learn spurious associations between signal morphology and sleep outcomes that do not transfer across populations. Robustness testing should therefore include stratified

evaluation across demographic groups, skin tones, clinical conditions, and device placements, rather than relying solely on aggregate accuracy metrics. System designers should also incorporate uncertainty estimation so that users and clinicians can distinguish high-confidence assessments from ambiguous or low-quality periods.

One structural approach to improving robustness is to treat sensor data as a multimodal ensemble rather than a single coherent stream. Accelerometry and photoplethysmography provide complementary information about sleep and wake states, and disagreement between modalities can serve as a signal of data quality problems or unusual physiological states. When the movement signal suggests wakefulness while cardiovascular signals suggest sleep, the system may flag the period for review rather than forcing a deterministic classification. This ensemble perspective aligns with clinical reasoning, which often weighs multiple sources of evidence rather than trusting a single indicator in isolation.

Robustness also depends on the frequency and granularity of model output. Continuous sleep quality assessment requires models to handle transitions between sleep stages, brief awakenings, and circadian variation. Models trained on discrete epoch-level labels may produce unstable predictions at boundaries or during fragmented sleep. Temporal smoothing can improve coherence but may obscure short arousals that have clinical importance. System architects must balance the desire for smooth, user-friendly sleep trajectories against the need to preserve clinically meaningful transient events. This is not merely a model tuning issue; it reflects a deeper interpretive question about what sleep quality means and how it should be represented to users and clinicians.

6. Fairness, Bias, and Clinical Generalizability

Fairness in AI-assisted sleep assessment requires attention to how wearable systems perform across different demographic, clinical, and social groups. Bias can enter through multiple pathways, including sensor design, training data composition, annotation procedures, and outcome definitions. If training datasets overrepresent younger, healthier, and lighter-skinned populations, models may perform poorly for older adults, individuals with chronic conditions, or people with darker skin tones. Algorithmic fairness research has demonstrated that predictive models in health care can encode and amplify existing disparities even when demographic variables are not explicitly included [13]. These risks are relevant to sleep AI because sleep disorders are unevenly distributed across populations and are often underdiagnosed in marginalized groups.

The measurement of sleep quality itself is not a neutral construct. Polysomnography-based labels reflect particular clinical definitions and scoring conventions that may not capture the full range of human sleep experience. When AI models are trained to predict these labels, they inherit their assumptions and limitations. For example, sleep stage scoring rules have historically been developed and validated in specific populations, and their application to diverse groups may introduce systematic errors. A model that reproduces polysomnography labels with high fidelity may still be unfair if the underlying labels are biased or if the model's errors are concentrated in particular subgroups. Fairness evaluation must therefore move beyond aggregate accuracy to examine subgroup-specific performance, error patterns, and clinical consequences [14].

Clinical generalizability is further complicated by the heterogeneity of wearable devices. Different manufacturers use different sensor configurations, firmware, and signal processing algorithms, which can change the relationship between raw signals and sleep outcomes. A

model trained on one device may not transfer directly to another, even if both devices report similar summary metrics. This device-level variability has implications for fairness because access to higher-quality devices may be unevenly distributed across socioeconomic groups. Health systems that rely on patient-owned wearables must therefore consider whether their AI pipelines can accommodate heterogeneous data sources without disadvantaging users who have older or less expensive devices.

Responsible development of sleep AI also requires attention to the broader determinants of sleep health. Sleep quality is shaped by housing conditions, occupational schedules, stress, environmental noise, and access to care. A wearable system that provides individualized sleep feedback without acknowledging these structural factors may place disproportionate responsibility on individuals while obscuring systemic contributors. Fairness-oriented design should incorporate contextual data cautiously, avoid stigmatizing language, and provide recommendations that are realistic given users' constraints. This sociotechnical perspective aligns with broader frameworks for responsible machine learning in health care, which emphasize that fairness is not a purely statistical property but an institutional commitment to equity [16].

7. Governance, Privacy, and Regulatory Considerations

Sleep data collected from wrist-worn sensors are highly sensitive because they can reveal daily routines, health status, mental state, and behavioral patterns. Governance frameworks must address data minimization, purpose limitation, user consent, and downstream data sharing. Unlike episodic clinical measurements, continuous sleep monitoring generates longitudinal records that may be re-identified even after direct identifiers are removed. Privacy risks are compounded when sleep data are linked with location, calendar, or smartphone usage data to infer contextual information. System architects should implement privacy-preserving practices such as on-device processing, differential privacy, and secure aggregation where feasible, while recognizing that these techniques introduce their own trade-offs in accuracy and computational cost.

Regulatory oversight of AI-assisted sleep assessment varies by jurisdiction and intended use. In the United States, software that provides diagnostic or treatment recommendations may be regulated as a medical device, whereas wellness-oriented sleep tracking may fall outside the Food and Drug Administration's traditional scope. The regulatory landscape is further complicated by the adaptive nature of machine learning systems, which may change after deployment through retraining or software updates. Regulatory agencies have proposed frameworks for managing modifications to artificial intelligence and machine learning-based software as a medical device, emphasizing the need for predetermined change control plans and ongoing performance monitoring [17]. These frameworks encourage manufacturers to anticipate model evolution and to document the safety and effectiveness implications of each update.

The European Union's approach to trustworthy artificial intelligence places additional emphasis on human oversight, transparency, and accountability [18]. Sleep AI systems that operate as decision support tools in clinical settings must provide clinicians with enough information to interpret model outputs and override them when necessary. This requirement has architectural implications because it favors systems that expose feature contributions, uncertainty estimates, and data quality indicators rather than returning opaque scores. Governance mechanisms should also include incident reporting channels, audit trails, and

mechanisms for users to contest decisions or correct erroneous data. These mechanisms require investment in organizational processes, not just technical components.

Ethical review of sleep AI should address secondary uses of data. Sleep data collected for personal wellness may later be repurposed for research, insurance underwriting, or employer wellness programs. Such repurposing can undermine user trust and create harms that extend beyond the original context. Data governance policies should clearly specify permissible uses, prohibit discriminatory applications, and provide users with meaningful control over their data. The mapping of ethical issues in health AI has highlighted the importance of transparency, responsibility, and the alignment of AI systems with patient values [15]. In sleep health, this means recognizing that sleep is simultaneously a biological process, a behavioral practice, and a social experience, and that AI systems should support rather than reduce this complexity.

8. Deployment and Operational Sustainability

Deploying an AI-assisted sleep assessment system in real-world settings requires more than a validated model. It demands an operational infrastructure capable of supporting device provisioning, user onboarding, data ingestion, model serving, report generation, and clinical follow-up. Many pilot systems rely on research coordinators and manual data cleaning, which is not sustainable at scale. Operational sustainability requires automated data quality monitoring, exception handling, and feedback loops that allow the system to improve over time without constant human intervention. At the same time, excessive automation can create new failure modes if users cannot reach human support when the system behaves unexpectedly.

Clinical integration is a critical deployment challenge. Sleep AI outputs must be interpretable by clinicians and actionable within existing workflows. If the system generates lengthy reports or uses unfamiliar sleep quality metrics, clinicians may ignore its recommendations or misinterpret its limitations. Successful integration often requires co-design with sleep medicine specialists, nursing staff, and patients to identify the most useful outputs and presentation formats. The clinical AI literature emphasizes that high-performing algorithms do not automatically produce clinical benefit; workflow integration, user trust, and decision support design are equally important [11]. In sleep medicine, AI tools are increasingly viewed as complements to clinical judgment rather than replacements for comprehensive evaluation [19].

Long-term sustainability also depends on economic and environmental factors. Continuous wearable monitoring generates substantial data volumes that require storage, processing, and transmission. Cloud-based architectures may incur recurring costs that are difficult to sustain in resource-constrained health systems. Edge computing can reduce data transfer costs but may require more frequent device replacement due to increased local processing demands. Environmental sustainability should also be considered, because the energy consumption of wearable devices, network infrastructure, and data centers contributes to the overall footprint of digital health systems. System designers should evaluate whether the incremental benefit of higher sampling rates or more frequent cloud synchronization justifies the additional energy and cost.

Another dimension of operational sustainability is the management of user expectations and behavior. Sleep AI systems often provide feedback intended to encourage healthier sleep behaviors, but repeated negative feedback can lead to anxiety or disengagement. The framing

of sleep quality results should account for natural night-to-night variability and avoid overmedicalizing normal sleep fluctuations. User-facing systems should include educational content that contextualizes sensor-derived metrics and distinguishes between clinically concerning patterns and ordinary variation. This human-centered design perspective is part of the broader governance responsibility of sleep AI developers and aligns with frameworks that prioritize doing no harm in machine learning for health [16].

Finally, ongoing performance monitoring is essential after deployment. Data distributions may shift as user populations change, sensor hardware is updated, or clinical definitions evolve. A model that performs well at launch may degrade silently over time if monitoring is not integrated into the operational pipeline. Performance monitoring should include accuracy against periodic reference data when available, as well as proxy indicators such as data quality scores, missingness rates, and subgroup performance. When degradation is detected, organizations need clear protocols for model recalibration, recall, or withdrawal. These operational practices are increasingly recognized as central to the responsible deployment of machine learning systems in health care [12]. Without them, the long-term safety and effectiveness of AI-assisted sleep assessment cannot be assured.

9. Conclusion

AI-assisted sleep quality assessment using wrist-worn physiological sensors represents a promising convergence of wearable sensing, machine learning, and sleep medicine. However, its value will be realized only if the supporting systems are designed with attention to architecture, robustness, fairness, governance, and operational sustainability. This paper has argued that sleep AI should be treated as a distributed socio-technical infrastructure rather than a standalone algorithmic achievement. The path from raw sensor data to actionable sleep insight involves a chain of decisions that shape privacy, accuracy, interpretability, and equity.

Structural trade-offs pervade every layer of the system. On-device processing protects privacy but limits model complexity. Cloud-based inference enables sophisticated analytics but increases data exposure and dependency on connectivity. Deep learning improves representational capacity but reduces interpretability. Population-level models enable broad deployment but may fail for underrepresented groups. These tensions cannot be fully eliminated, but they can be managed through explicit design choices, transparent governance, and continuous monitoring. The development of robust wrist-worn vital sign extraction under varying conditions is a foundational capability that supports these broader system goals, but it must be embedded within an architecture that also addresses quality control, data provenance, and clinical integration.

Future work should move beyond single-site validation toward multi-device, multi-population evaluations that test generalizability under realistic conditions. It should also investigate how sleep AI systems can support health equity rather than reinforce existing disparities. This will require collaboration among engineers, clinicians, epidemiologists, ethicists, and policymakers to define appropriate performance criteria, governance standards, and deployment models. As wearable sensors become more ubiquitous and AI methods more powerful, the central challenge will not be whether sleep stages can be predicted from wrist signals, but whether the resulting systems can be trusted, sustained, and made beneficial across diverse populations and contexts. The academic and clinical communities must therefore approach AI-assisted sleep assessment with both technical rigor and systemic responsibility.

References

1. Perez-Pozuelo, I., Zhai, B., Palotti, J., Mallol, R., Aupetit, M., Garcia-Gancedo, L., ... & Flach, P. (2020). The future of sleep health: A data-driven revolution in sleep science and medicine. NPJ Digital Medicine, 3, 42.

2. de Zambotti, M., Cellini, N., Goldstone, A., Colrain, I. M., & Baker, F. C. (2019). Wearable sleep technology in clinical and research settings. Medicine & Science in Sports & Exercise, 51(7), 1538–1557.

3. Sadeh, A., Sharkey, K. M., & Carskadon, M. A. (1994). Activity-based sleep-wake identification: An empirical test of methodological issues. Sleep, 17(3), 201–207.

4. van Hees, V. T., Sabia, S., Anderson, K. N., Denton, S. J., Oliver, J., Catt, M., ... & Singh-Manoux, A. (2015). A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PLoS ONE, 10(11), e0142533.

5. Walch, O., Huang, Y., Forger, D., & Goldstein, C. (2019). Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. Sleep, 42(12), zsz180.

6. Haghayegh, S., Khoshnevis, S., Smolensky, M. H., Diller, K. R., & Castriotta, R. J. (2019). Accuracy of wristband Fitbit models in assessing sleep: Systematic review and meta-analysis. Journal of Medical Internet Research, 21(11), e16273.

7. Radha, M., Fonseca, P., Moreau, A., Ross, M., Cerny, A., Anderer, P., ... & Aarts, R. M. (2021). A deep transfer learning approach for wearable sleep stage classification with photoplethysmography. NPJ Digital Medicine, 4, 135.

8. Fonseca, P., van Gilst, M. M., Radha, M., Ross, M., Moreau, A., Cerny, A., ... & Overeem, S. (2020). Automatic sleep staging using heart rate variability, body movements, and recurrent neural networks in a sleep disordered population. Sleep, 43(9), zsaa048.

9. Chen, M., Ma, Y., Song, J., Hu, B., & Wang, Y. (2017). Internet of Things for smart healthcare: Technologies, challenges, and opportunities. IEEE Access, 5, 26521–26544.

10. 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.

11. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.

12. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.

13. 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.

14. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.

15. Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. Social Science & Medicine, 260, 113172.

16. 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.

17. U.S. Food and Drug Administration. (2019). Proposed regulatory framework for modifications to artificial intelligence/machine learning-based software as a medical device. FDA.

18. European Commission. (2019). Ethics guidelines for trustworthy AI. Publications Office of the European Union.

19. Bandyopadhyay, A., & Goldstein, C. (2023). Clinical applications of artificial intelligence in sleep medicine: A sleep clinician's perspective. Sleep and Breathing, 27(1), 39–50.

20. Kolla, B. P., Mansukhani, S., & Mansukhani, M. P. (2016). Consumer sleep tracking devices: A review of mechanisms, validity and utility. Expert Review of Medical Devices, 13(5), 497–506.