

Personalized Machine Learning for Exercise Recovery Assessment Using Continuous Wearable Sensor Data

Dominik A. Schwartz

Department of Computer Science, University of North Texas, Denton, TX, USA.
schwartz-dominik@unt.edu

Meiqiu Guo

Department of Computer Science, University of Central Florida, Orlando, FL, USA.
meiqiuguo1996@ucf.edu

Leon L. Dawson

Department of Computer Science, University of Houston, Houston, TX, USA.
lldawson15@uh.edu

Finn Cox

Department of Electrical Engineering and Computer Science, University of Missouri,
Columbia, MO, USA.
finn.dev@missouri.edu

Abstract

Exercise recovery assessment is a central but under-instrumented component of health management, athletic training, and clinical rehabilitation. Subjective readiness questionnaires and infrequent laboratory tests cannot capture the continuous physiological dynamics that determine whether an individual is adequately recovered for subsequent exertion. Wearable sensors now offer an unprecedented opportunity to monitor movement, cardiac activity, respiration, sleep, and related signals over long periods. Personalized machine learning can transform these noisy, heterogeneous, and contextual data streams into meaningful recovery estimates. However, translating algorithmic potential into dependable real-world infrastructure raises a distinct set of systems challenges. This paper presents a cross-disciplinary analysis of personalized machine learning for exercise recovery assessment as a socio-technical infrastructure problem. It examines sensor data acquisition, edge and cloud architectural trade-offs, personalization strategies, model interpretability, robustness, fairness, governance, and long-term sustainability. Rather than proposing a narrow model architecture, the paper argues that effective recovery assessment requires an integrated pipeline in which physiological signal processing, adaptive learning, privacy-preserving aggregation, and human-centered oversight are jointly designed. Structural trade-offs include the balance between local low-latency inference and global model quality, between individual adaptation and population-level generalization, and between rich physiological modeling and energy-efficient deployment. The discussion further addresses fairness risks arising from sensor heterogeneity and demographic underrepresentation, as well as the governance requirements for safe clinical and consumer use. The paper concludes that sustainable personalized recovery systems will depend less on isolated algorithmic gains than on the deliberate

alignment of data infrastructure, model lifecycle management, regulatory accountability, and equitable access.

Keywords

personalized machine learning; wearable sensors; exercise recovery; digital health; system architecture; federated learning; algorithmic fairness; governance; sustainability.

1. Introduction

Machine learning is increasingly positioned as a transformative force in health and wellness applications because it can identify complex patterns in data that traditional threshold-based rules fail to capture [1]. This transformation has been particularly visible in high-performance medicine, where multimodal physiological data are expected to support earlier and more individualized decisions [2]. At the same time, the widespread adoption of consumer wearable devices has created a continuous observational layer around everyday behavior, including exercise, mobility, sleep, and autonomic regulation [3]. These devices generate dense longitudinal signals that are highly relevant for exercise recovery assessment because recovery is not a single event but a prolonged, multidimensional process influenced by prior training load, sleep quality, nutrition, stress, hydration, and illness.

Despite the abundant data, converting wearable streams into reliable recovery states is not a straightforward pattern recognition task. Recovery status is a latent construct that cannot be directly measured by one sensor, and its expression varies substantially across individuals. The same heart rate variability pattern may indicate adequate recovery for one person and accumulating fatigue for another. Wearable data also inherit significant noise from motion artifacts, sensor placement differences, skin type variation, and device firmware heterogeneity. These challenges call for personalized machine learning rather than population-averaged scoring. However, personalization introduces further complexity because individual models require sufficient longitudinal baselines, must adapt to nonstationary physiology, and must avoid overfitting to spurious daily fluctuations [4]. The machine learning system must therefore be understood as an infrastructure that includes data acquisition, preprocessing, model training, inference, monitoring, and human interpretation.

This paper adopts a system-level perspective on personalized machine learning for exercise recovery assessment. It argues that the field has matured beyond simple feature engineering and must now confront structural trade-offs related to edge computing, cloud orchestration, data governance, model drift, fairness, and sustainability. The analysis emphasizes that technical design choices are inseparable from organizational and policy choices. A recovery model may perform well in a controlled study but fail in deployment if it cannot tolerate missing sensor data, inconsistent user behavior, demographic heterogeneity, or limited connectivity. Therefore, the following sections examine the architecture, personalization strategies, statistical considerations, robustness, fairness, deployment governance, and long-term maintenance of such systems. The objective is not to prescribe a single solution but to develop a coherent conceptual framework for designing, deploying, and governing personalized recovery assessment infrastructure [5].

2. Background and Related Work

Wearable data analytics have been characterized as windows into human health because they reveal physiological dynamics that are otherwise invisible to clinicians and individuals between clinical encounters [6]. For exercise recovery, relevant sensor channels include

accelerometry, photoplethysmography, electrocardiography, skin temperature, and in some cases electrodermal activity. From these channels, derived indicators such as resting heart rate, heart rate variability, sleep duration and architecture, respiratory rate, and movement quality can be extracted. Machine learning has been applied broadly to interpret such signals for health applications, with supervised, unsupervised, and semisupervised methods being used to detect abnormal states, predict outcomes, and cluster behavioral patterns [7]. These methods are attractive because they can integrate many weak signals into a robust latent estimate, but their success depends heavily on data quality and contextual alignment.

Federated learning has emerged as a significant privacy-preserving paradigm for healthcare informatics because it allows model updates to be computed on local devices or local servers without centralizing raw physiological data [8]. This is particularly relevant for recovery assessment, where continuous heart rate, activity, and sleep data are sensitive. In parallel, deep learning methods developed for electronic health records have demonstrated how sequential clinical data can be modeled through architectures that capture long-range dependencies and irregular sampling [9]. Such approaches have influenced wearable analytics by encouraging the treatment of physiological time series as longitudinal records rather than independent snapshots. However, many clinical deep learning models require curated datasets, whereas wearable recovery applications must operate in uncontrolled environments with substantial missingness and noise.

A closely related line of work addresses the challenge of physiological monitoring during physical activity, when motion artifacts severely degrade sensor signals. A joint attention mechanism learning approach has been proposed to facilitate opto-physiological monitoring during physical activity [10]. This line of research is relevant because recovery assessment depends on accurate physiological readings obtained during and after exercise, where movement-related corruption is most severe. Attention mechanisms can help the model focus on informative segments and learn to reject motion-distorted intervals. Although the present paper does not center on any single signal processing method, such advances illustrate the broader shift toward learned denoising and context-aware physiological inference.

3. System Architecture and Data Infrastructure

The system architecture for personalized exercise recovery assessment can be understood as a layered structure spanning wearable sensors, on-device processing, edge gateways, cloud services, and organizational data infrastructure. On-device processing is necessary for low-latency feature extraction, artifact detection, and immediate user feedback, while cloud infrastructure supports longitudinal model training, cross-user aggregation, and computationally intensive analysis. Edge computing offers an intermediate point for processing data close to the data source, reducing communication load and enabling timely feedback even when bandwidth is constrained [11]. The placement of computation across these layers is not merely a technical optimization; it determines privacy properties, energy consumption, update frequency, and the degree to which personalization can operate locally.

Data infrastructure must accommodate continuous high-frequency streams that vary in completeness and semantic consistency. A well-designed infrastructure includes ingestion pipelines that validate device clocks, handle firmware differences, and align sensor channels to common timestamps. Data quality controls should detect implausible values, repeated values, and sensor disconnection, while also preserving raw segments that may be useful for later reprocessing. Large-scale wearable studies have shown that analysis decisions such as nonwear detection, filtering, and aggregation windows can substantially change downstream

conclusions [4]. These decisions are particularly consequential for recovery assessment because recovery-related signals such as nocturnal heart rate variability require stable sleep periods and minimal movement artifacts.

The architecture also needs to support multiple stakeholders with different information needs. Athletes may require actionable recovery guidance, coaches may need aggregated squad readiness dashboards, and clinicians may need explainable summaries for rehabilitation decisions. This heterogeneity argues against a single monolithic model and in favor of a service-oriented design in which shared data assets and feature stores feed multiple model endpoints. Interoperability with electronic health records, training logs, and nutrition applications further increases complexity. Consequently, the data infrastructure should be treated as a core research and engineering object rather than as a passive input to machine learning [5]. The structural trade-off between storing rich raw data for future model revision and respecting storage capacity, privacy rules, and data minimization principles is especially important.

4. Personalization Strategies for Recovery Modeling

Recovery is a multidimensional construct that cannot be reduced to a single physiological marker. Effective personalization requires modeling the relationships among training load, autonomic state, sleep, subjective symptoms, and environmental context within an individual over time [6]. Rather than applying a fixed threshold to heart rate variability or resting heart rate, a personalized system learns each individual baseline and the expected deviation following different types of exertion. This baseline must be allowed to drift slowly as fitness, illness, medication, and lifestyle change. If the baseline is too rigid, the model will generate false alarms during ordinary physiological adaptation; if it is too flexible, the model will normalize dangerous fatigue or overreaching.

A range of model families can support personalized recovery inference. Recurrent neural networks, particularly long short-term memory networks, are well suited to sequential physiological data because they can maintain internal state across time and learn temporal dependencies [12]. Attention-based architectures provide another compelling option because they allow the model to weight different time segments and sensor channels according to their relevance for a given prediction [13]. These deep learning approaches can capture complex temporal interactions, but they require substantial data and careful regularization. Tree-based ensemble methods such as gradient boosting remain valuable for tabular features derived from wearable signals, especially when training data are modest and model interpretability is important [14]. In practice, a robust personalized system may combine gradient boosting for tabular features with recurrent or attention models for raw or lightly processed sequence data.

Transfer learning offers a powerful mechanism for personalization without requiring each new user to provide months of labeled data. A base model can be trained on a large population and then adapted to a new individual through fine-tuning or domain adaptation [16]. This approach exploits shared physiological structure while allowing individual deviations to emerge. However, transfer learning can also import population-level biases into the personalized model, especially when the source population differs demographically from the target user. Before deployable models are trusted, post hoc explanation methods such as Shapley additive explanations can help clarify which features drive an individual recovery prediction [15]. These explanations are not complete causal accounts, but they provide an important safeguard against hidden failure modes and support clinician and user trust.

Personalization also interacts with privacy architecture. Federated personalization can maintain a global model across users while learning local adapters that remain on the user device or within a trusted local server [8]. This arrangement preserves the benefits of large-scale learning while reducing the risk associated with transferring identifiable physiological time series. The central systems challenge is to balance the communication and synchronization costs of federated training with the need for responsive adaptation to individual patterns. A purely local model may adapt quickly but learns nothing from population data, whereas a purely centralized model may generalize well but personalizes slowly and raises privacy concerns.

5. Computational and Statistical Considerations

Developing personalized recovery models requires careful attention to statistical validity, temporal ordering, and distributional shift. Wearable data are opportunistic and unevenly sampled, meaning that missingness is rarely fully random. Users may wear devices less during illness or high-stress periods, creating systematic gaps precisely when recovery signals are most clinically important. Data leakage is another pervasive risk when future observations, labels, or post hoc aggregations are inadvertently included in training features [4]. For recovery assessment, leakage can arise when sleep metrics from the entire night are used to predict morning readiness at a time before the full night is known, or when training labels are derived from subsequent performance without respecting temporal boundaries.

Model evaluation must therefore reflect the intended deployment condition. Random splits across timestamps are often overly optimistic because they allow the model to learn individual patterns from nearby periods and then test on temporally adjacent data. Blocked validation by user and by calendar period provides a stronger test of both personalization and generalization. The evaluation should also report performance separately across demographic groups, device types, and activity contexts to reveal hidden failures. Responsible machine learning frameworks emphasize that aggregate performance can mask clinically meaningful disparities [17]. A recovery model may appear accurate overall while systematically underestimating recovery for users with higher melanin skin tones if photoplethysmography signals are less reliable for those users.

Computational complexity is often neglected in the machine learning literature but is decisive in wearable deployment. Continuous inference on a constrained device can drain battery and reduce sensor sampling quality. Models that require large attention computations or deep recurrent processing may need to be distilled or converted to efficient approximations before on-device use. Cloud-based inference avoids local energy constraints but depends on connectivity and increases privacy exposure. A hybrid approach can use a lightweight on-device model for immediate feedback and a more expressive cloud model for overnight or asynchronous analysis. This division of labor is itself a statistical trade-off because the cheap model may be less accurate and its errors may shape subsequent user behavior before the more accurate cloud model can intervene.

6. Robustness, Fairness, and Generalization

Robustness is a first-order requirement for recovery assessment because wearable data are collected in uncontrolled environments. Sensor placement, skin contact pressure, ambient temperature, and motion intensity can all alter the raw signal. If the model learns to depend on features that are unstable across devices or firmware versions, its performance may degrade when users upgrade their hardware. Robustness can be improved through data augmentation,

multi-device harmonization, and invariant feature learning, but it cannot be guaranteed by algorithmic means alone. Deployment pipelines must continuously monitor data quality and model confidence, flagging predictions that fall outside the support of the training distribution.

Fairness concerns arise from both representation and measurement. Training datasets for wearable analytics are often skewed toward younger, healthier, higher-income, and more physically active populations. Algorithms trained on such data may produce less accurate recovery estimates for older adults, individuals with cardiovascular conditions, or users from underrepresented demographic groups [18]. Measurement bias compounds this problem because photoplethysmography and accelerometry can perform differently across skin tones, body compositions, and movement patterns. A widely discussed study in healthcare showed that an algorithm can systematically disadvantage a racial group even when race is not explicitly included [19]. In exercise recovery, similar mechanisms could lead a model to understate fatigue for certain groups, resulting in unsafe training recommendations or delayed clinical attention.

Generalization across user populations and contexts must be explicitly designed rather than assumed. A model validated in a controlled laboratory study with high compliance may fail in free-living conditions where users do not wear sensors consistently or where sleep schedules are irregular. Dataset shift can occur gradually as seasons change, training routines shift, or health status evolves. Continuous monitoring of prediction distributions and periodic recalibration are therefore essential. Transfer learning and domain adaptation can improve generalization, but they also introduce risks if source and target distributions are poorly aligned [16]. Fairness and generalization should be treated as ongoing system properties rather than one-time certification tasks.

7. Deployment, Governance, and Policy Implications

Deploying personalized recovery assessment systems requires governance that extends beyond conventional software quality assurance. These systems influence health-related decisions, and their errors can lead to overtraining, missed illness, or inappropriate reassurance. A responsible machine learning roadmap for health care emphasizes the need for prospective validation, clear task definition, and human accountability [17]. In consumer wellness contexts, regulatory expectations may be lower, but the potential harms remain significant because users often interpret recovery scores as medical-grade guidance. Organizations should therefore define whether the system is a wellness aid, a clinical decision support tool, or a regulated medical device, because this classification shapes validation, labeling, and legal liability.

Data governance is especially important because continuous physiological data can reveal sensitive information about sleep, stress, cardiovascular function, and daily routines. Federated learning and differential privacy can reduce some privacy risks, but they do not eliminate governance responsibilities [20]. Users must have meaningful consent processes that explain how their data will be used, how long it will be retained, and which third parties may access model outputs. Interoperability standards are needed to ensure that recovery assessments can be integrated into electronic health records without creating fragmented data silos. Without clear data stewardship policies, personalized recovery models may entrench power asymmetries in which platforms hold rich longitudinal data while users receive only opaque scores.

The governance framework should also address algorithmic change over time. A recovery model that is updated weekly or daily is not the same product from one month to the next. Model versioning, audit trails, and rollback procedures are necessary for accountability. When a model changes, users and clinicians should be able to understand what changed and why predictions shifted. This requires not only technical monitoring but also organizational review processes that include clinical, ethical, and user perspectives. Clinical integration further demands workflows that do not overwhelm providers with alerts. Recovery scores should be embedded in shared decision-making rather than used as independent prescriptions [2]. Ultimately, governance should create the conditions under which personalized models are safe, contestable, and continuously improved.

8. Sustainability and Long-Term Maintenance

Sustainability is a critical but often overlooked dimension of personalized recovery infrastructure. The resource footprint of continuous wearable data collection, cloud storage, model retraining, and edge inference can be substantial. Large-scale health data projects incur not only financial costs but also environmental costs associated with computation and storage [4]. A sustainable system should avoid unnecessary data retention, use energy-efficient inference paths, and compress raw data into informative features when possible. However, data minimization must be balanced against the need to revisit raw signals when new algorithms or physiological insights emerge. This balance requires a lifecycle approach in which data value, privacy risk, storage cost, and environmental impact are periodically reassessed.

Long-term maintenance also includes organizational capacity and workforce expertise. Recovery models are not static artifacts; they require continuous monitoring, retraining, and evaluation as sensor hardware evolves and user populations change. Federated learning offers one pathway for sustainable model updates, because local computation reduces centralized data movement and can support ongoing improvement without continuously transferring raw sensor data [20]. However, federated systems introduce their own sustainability challenges, including heterogeneity of client devices, communication scheduling, and the need for secure aggregation at scale. A federated deployment may reduce privacy burden while increasing coordination complexity and requiring stronger engineering infrastructure.

The long-term success of personalized recovery systems depends on aligning incentives among users, device manufacturers, clinicians, platform operators, and regulators. If the system is optimized solely for engagement, it may produce frequent but unstable predictions that encourage compulsive checking rather than meaningful recovery behavior. If it is optimized solely for clinical caution, it may issue so many warnings that users disengage. Sustainable design therefore includes human factors, behavioral theory, and shared governance, not only machine learning. Future research should examine how recovery models can support long-term behavior change without increasing health anxiety, and how community-based or clinical deployment models can deliver equitable access beyond commercial wellness markets [5]. The most sophisticated algorithm will fail if the surrounding infrastructure cannot maintain trust, reliability, and relevance over years of changing user behavior and technology.

9. Conclusion

Personalized machine learning for exercise recovery assessment represents a convergence of wearable sensing, longitudinal physiological modeling, edge computing, and responsible data

governance. This paper has argued that the central challenge is not simply to build a more accurate model but to design a system that is robust under real-world noise, fair across diverse populations, interpretable enough for human accountability, and sustainable over time. Continuous wearable data offer an unprecedented view into individual recovery dynamics, but they also create infrastructural burdens related to data quality, privacy, model drift, and operational governance.

The structural analysis presented here shows that personalization must be approached as a layered and adaptive process rather than a fixed model choice. Local on-device inference, edge preprocessing, cloud aggregation, and federated adaptation each offer different trade-offs in latency, energy, privacy, and generalization. These trade-offs cannot be resolved through algorithmic innovation alone; they require deliberate architectural decisions and governance mechanisms that reflect the intended use context. Fairness and robustness must be treated as ongoing properties, monitored continuously rather than checked once before launch.

Future work should integrate causal reasoning, multimodal contextual data, and participant-centered governance into recovery assessment systems. It should also investigate how personalized recovery models can be validated across clinical, athletic, and general wellness settings without overstating their predictive certainty. The long-term value of these systems will depend on their ability to respect user agency, reduce rather than amplify inequities, and support safe, evidence-informed decisions about recovery and exertion.

References

1. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29.
2. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
3. Piwek, L., Ellis, D. A., Andrews, S., & Joinson, A. (2016). The rise of consumer health wearables: Promises and barriers. *PLoS Medicine*, 13(2), e1001953.
4. Hicks, J. L., Althoff, T., Sosic, R., Kuhar, P., Bostjancic, B., King, A. C., Leskovec, J., & Delp, S. L. (2019). Best practices for analyzing large-scale health data from wearables and smartphone apps. *npj Digital Medicine*, 2, 45.
5. Smuck, M., Odonkor, C. A., Wilt, J. K., Schmidt, N., & Swiernik, M. A. (2021). The emerging clinical role of wearables: Factors for successful implementation. *npj Digital Medicine*, 4, 45.
6. 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.
7. Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2, 160.
8. Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2020). Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research*, 5, 1–19.
9. Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record data. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1589–1604.

10. Zheng, X., Hu, S., Dwyer, V., Barrett, L., & Derakhshani, M. (2026). Joint attention mechanism learning to facilitate opto-physiological monitoring during physical activity. *Biomedical Signal Processing and Control*, 113, 108949.
11. Muhammad, G., Alhamid, M. F., & Long, X. (2018). Computing and processing on the edge for smart health. *IEEE Communications Magazine*, 56(11), 96–101.
12. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.
14. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
15. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.
16. Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems*, 32.
17. Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Goldenberg, A., Kehl, K., Mooney, S., Moses, O., Naumann, T., Ong, M., Rosing, T., Schulam, P., Sittig, D. F., Thadaney, S., Vayena, E., & Vickers, A. J. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine*, 25(9), 1337–1340.
18. 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), 115.
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. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *npj Digital Medicine*, 3, 119.