Machine Learning-Based Fatigue Assessment Using Continuous Physiological and Motion Signals

Yuanke Jia
Department of Computer Science, George Mason University, Fairfax, VA, USA.
jia770@gmu.edu

Dennis Green
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
dgreen@unr.edu

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

Quanxi Weng
Department of Computer Science, University of Houston, Houston, TX, USA.
quanxiweng01@uh.edu

Abstract

Fatigue is a multidimensional state with physiological, cognitive, and behavioral components that can impair human performance and increase the likelihood of errors in safety-critical environments. Continuous wearable sensing now enables the collection of physiological and motion signals outside controlled laboratory settings, creating new opportunities for machine learning-based fatigue assessment. However, translating raw sensor streams into reliable, fair, and actionable fatigue predictions requires more than algorithmic innovation. This paper presents a systems-level analysis of fatigue assessment infrastructure, integrating perspectives from sensing, machine learning, deployment engineering, governance, and policy. The discussion identifies structural trade-offs in edge, cloud, and federated architectures, examines data processing and signal quality challenges, and compares feature-based and representation learning approaches. It further considers robustness under real-world variability, fairness across population groups, and accountability mechanisms for high-stakes applications. The paper argues that fatigue assessment systems should be designed as governed socio-technical infrastructures rather than isolated classification tools. Long-term sustainability depends on data provenance, drift monitoring, human oversight, and alignment with regulatory and labor protections. The analysis offers a conceptual framework for researchers and system designers seeking to build fatigue assessment systems that are not only accurate but also trustworthy, interpretable, and institutionally responsible.

Keywords

fatigue assessment; machine learning; wearable sensing; physiological signals; system architecture; fairness; real-time monitoring.

1. Introduction

Fatigue is a multidimensional construct involving physiological, cognitive, and behavioral components. In operational environments such as transportation, industrial work, and health care, fatigue can degrade attention, slow reaction time, reduce situation awareness, and increase the risk of accidents. Traditional methods for measuring fatigue have relied on subjective self-report scales, periodic performance tests, or coarse scheduling heuristics. These methods are useful in controlled settings but cannot capture the dynamic and individualized nature of fatigue as it develops over time. The growth of wearable sensing technologies has created new possibilities for continuous observation of physiological and motion signals, including heart rate, heart rate variability, electrodermal activity, skin temperature, and accelerometer-derived movement. Early work on personal sensing emphasized the value of longitudinal mobile and wearable data for understanding psychological and behavioral states [1]. This line of research established that continuous signal collection, when paired with appropriate analytical methods, can reveal patterns that are not visible through episodic assessments.

Wearable health monitoring has evolved from simple step counting to sophisticated multimodal sensing. Data mining approaches for wearable sensor platforms have demonstrated that movement and physiological data can support health-related inference, but they also reveal persistent challenges such as missing data, sensor noise, annotation burden, and population heterogeneity [2]. Benchmark datasets for activity monitoring have enabled more reproducible comparisons of motion-based machine learning models, particularly for recognizing activities that are relevant to fatigue context [3]. Reviews of wearable sensing systems further emphasize that the utility of such platforms depends less on any single sensor or algorithm and more on the integrated design of sensing, processing, and decision-making layers [4]. These observations support a systems-level perspective: fatigue assessment should be treated as an infrastructure problem rather than a narrow pattern recognition task.

This article examines machine learning-based fatigue assessment using continuous physiological and motion signals from an interdisciplinary systems perspective. It considers architectural trade-offs, signal processing infrastructure, modeling choices, deployment conditions, robustness, fairness, and policy implications. Rather than proposing a single algorithmic solution, the paper develops a conceptual framework for designing and governing fatigue assessment systems that can operate reliably at scale. The discussion draws on developments in wearable sensing, machine learning, and socio-technical governance to identify structural tensions and forward-looking directions.

2. System Architecture for Multimodal Fatigue Assessment

A functional fatigue assessment system comprises several interacting layers: sensing, local preprocessing, communication, data aggregation, learning, inferencing, and feedback. The choice of where to place computational workload across wearable devices, edge gateways, and cloud services is a central architectural decision. On-device processing offers low latency and can operate when connectivity is intermittent, but it is constrained by battery capacity, thermal dissipation, memory, and processor limitations. Cloud processing enables larger models and longitudinal analysis, but it introduces latency and depends on reliable network connectivity. Edge gateways provide an intermediate option, aggregating data from multiple wearables and executing models with moderate complexity. The architecture must balance these trade-offs according to the operational context, the type of fatigue being assessed, and the safety requirements of the deployment environment. Continuous fatigue assessment in transportation or industrial settings may require near-real-time alerts, whereas clinical or

wellness applications may tolerate delayed analysis. In all cases, the system must accommodate motion artifacts and changes in physical activity intensity. Rapid extraction of vital signs from opto-physiological signals under varying activity intensity is now technically feasible and provides an important building block for such systems [5]. This capability supports the integration of physiological monitoring with motion signals in unconstrained settings.

A well-designed architecture also separates data acquisition from analytical logic to allow independent evolution of sensors and models. The sensing layer should produce standardized data streams that include timestamps, signal quality indicators, and metadata about device placement and environmental context. The data plane must handle buffering, compression, and synchronization because physiological signals such as electrocardiography and photoplethysmography operate at different sampling rates than accelerometry and gyroscope signals. Without careful temporal alignment, multimodal models may learn spurious relationships. The learning plane consumes curated feature streams and supports training, personalization, and continual updating. The decision plane converts model outputs into actionable feedback, which may be presented to the individual, a supervisor, or an automated control system. The separation between decision and learning layers is important for liability and governance, because a change in model version should not necessarily trigger immediate operational changes without validation. This layered perspective is common in large-scale digital infrastructure but often underappreciated in wearable health applications.

Architectural choices also affect privacy and data ownership. When raw physiological data are transmitted to central servers, the risk of re-identification and secondary use increases. Conversely, exclusive on-device processing may limit the ability to conduct population-level fairness audits or model refinement. Federated architectures may provide a compromise by allowing model updates to be shared without transferring raw signals, although such approaches introduce coordination complexity and may still leak information through model parameters. The architecture must therefore be evaluated not only for latency, accuracy, and energy efficiency, but also for its implications for consent, transparency, and accountability. Fatigue assessment systems are embedded in employment, transportation, and health care contexts, where data misuse or erroneous predictions can have serious consequences. These structural considerations shape the reliability and social acceptability of the technology.

3. Signal Processing and Data Infrastructure

Physiological signals carry information about autonomic nervous system activity, cardiovascular function, and thermoregulatory processes that are relevant to fatigue. Photoplethysmography is a widely used optical technique for measuring blood volume pulse and can support heart rate and oxygen saturation estimation in wearable devices [6]. However, photoplethysmographic signals are highly sensitive to motion artifact, skin tone, tissue compression, and ambient light. Pulse rate variability derived from photoplethysmography can approximate heart rate variability but differs in timing accuracy and artifact susceptibility [7]. These limitations require careful preprocessing, signal quality assessment, and context awareness. In driving studies, physiological sensors have been used to detect stress and arousal under real-world conditions, demonstrating both the promise and the practical difficulty of collecting clean physiological data outside the laboratory [8]. Fatigue assessment systems must therefore incorporate quality filtering, adaptive signal reconstruction, and motion context modeling before machine learning models are applied.

The processing pipeline typically includes synchronization, artifact removal, segmentation, feature extraction, and normalization. For continuous use, the system should compute features over overlapping windows, but the optimal window length depends on the physiological process and the operational decision horizon. Short windows of a few seconds may capture rapid changes in heart rate but are vulnerable to high-frequency noise, while longer windows of several minutes may smooth out important transient responses. Motion signals can serve both as contextual inputs and as reference signals for artifact suppression. For example, accelerometer patterns can indicate walking, running, or vehicle vibration, thereby helping the system distinguish true physiological changes from movement-induced corruption. The data infrastructure must record not only the processed features but also uncertainty estimates, because model predictions made on low-quality segments require different decision thresholds.

Scalable data infrastructure also depends on semantic consistency across devices, sensors, and sites. Differences in sampling rates, device firmware, sensor placement, and calibration can create distribution shifts that degrade model performance. The system should capture metadata about device model, sensor configuration, body location, and environmental conditions to support robust modeling. Data versioning, lineage tracking, and provenance management are necessary for auditing and for understanding how training data differ from deployment data. In many practical settings, fatigue labels are scarce, noisy, or subjective. The data infrastructure must support multiple label sources, including self-report, task performance, behavioral indicators, and operational outcomes. The integration of these heterogeneous sources requires careful ontology design and governance, because inconsistent labels can propagate bias and reduce model credibility. Without investment in the data layer, advanced machine learning methods cannot deliver reliable fatigue assessment in real environments.

4. Machine Learning Modeling Approaches and Structural Trade-offs

Machine learning models for fatigue assessment can be grouped into classical feature-based approaches and representation learning approaches. Gradient-boosted decision trees remain effective for tabular features derived from physiological and motion signals, offering strong performance with modest computational requirements and relative interpretability compared with deep models [9]. Recurrent neural networks, particularly long short-term memory architectures, have been used to capture temporal dependencies in physiological sequences, allowing models to learn patterns of progressive fatigue over minutes or hours [10]. Attention mechanisms further enable models to focus on relevant time segments and sensor modalities, improving flexibility for multimodal data [11]. The choice among these methods involves trade-offs among predictive performance, training data requirements, latency, energy consumption, explainability, and ease of deployment.

Feature-based models often require domain-informed feature engineering, including heart rate variability indices, motion intensity metrics, and circadian context. This engineering can be labor-intensive but supports regulatory review and clinical interpretation. Deep learning models can learn representations directly from raw or lightly processed signals, but they generally require larger datasets and can be sensitive to artifacts and distribution shifts. Multimodal deep learning may fuse physiological and motion streams at early, intermediate, or late stages, and the fusion architecture affects robustness and interpretability. A shared representation may improve predictive power but can obscure the contribution of each modality, making debugging and fairness analysis more difficult. In contrast, late fusion

allows separate evaluation of modalities but may miss nonlinear interactions. These design decisions are not purely technical; they shape how failures are understood and attributed.

Training fatigue models at scale raises data governance and computational concerns. Centralized training on pooled user data may improve model generalization but creates privacy risks and may conflict with data protection regulations. Federated learning offers a distributed alternative where model updates are computed locally and aggregated without sharing raw physiological signals [12]. However, federated systems face challenges such as non-identically distributed data, client availability, communication efficiency, and potential leakage through gradient updates. The choice between centralized, federated, and fully on-device training must consider the sensitivity of health data, the need for model auditing, and the feasibility of maintaining model quality over time. Fatigue assessment systems also require personalization, because baseline physiology and fatigue responses vary across individuals. Personalization can be achieved through fine-tuning or through user-specific calibration, but this introduces additional model management complexity. The system must track model versions, personalization state, and deployment context to ensure consistent behavior.

5. Deployment, Robustness, and Real-Time Governance

Deployment of fatigue assessment systems in operational environments demands high reliability under variable conditions. Sensor detachment, battery depletion, network interruption, and environmental extremes can degrade data quality and model availability. The system should include health monitoring for the sensing infrastructure itself, producing alerts when data quality falls below acceptable thresholds. Graceful degradation is essential: when physiological signals are unavailable, the system may rely on motion patterns or task-based indicators, but it should communicate reduced confidence. Decision-making based on machine learning predictions should be calibrated to the consequences of false positives and false negatives. In safety-critical applications, false negatives may be more dangerous, while in wellness applications, false positives may reduce user trust. The deployment architecture must allow thresholds to be adjusted according to operational policy without requiring model retraining.

Real-time governance refers to the policies, controls, and oversight mechanisms that surround an automated fatigue assessment pipeline. Continuous monitoring can easily become surveillance if data are used beyond their original purpose. The political economy of large-scale sensing raises concerns about worker control, behavioral nudging, and the commodification of health data [13]. Fatigue prediction should not become a tool for intensifying work or penalizing employees. Governance mechanisms such as data minimization, purpose limitation, user access, and independent auditing are necessary to align the technology with social values. In transportation and industrial contexts, algorithmic recommendations may be fused with human supervisory judgment rather than automatic enforcement. This human-in-the-loop approach preserves accountability and allows contextual knowledge to influence decisions.

Robustness also extends to model performance across demographic and physiological groups. Evidence from health care algorithms has shown that predictive models can systematically disadvantage certain populations when training data and outcome proxies encode structural inequalities [14]. Fatigue assessment models trained on convenience samples may perform poorly for older adults, shift workers, people with chronic conditions, or individuals with different skin pigmentation, especially when optical sensors are used. Fairness cannot be

achieved through a single metric; it requires attention to data representativeness, label bias, group-specific performance, and differential consequences of errors [15]. Systems must be evaluated not only by aggregate accuracy but also by examining failure modes across subgroups and operational contexts. Fairness audits should be conducted before deployment and periodically after deployment as populations and conditions change.

6. Policy, Accountability, and Sustainability

The governance of machine learning-based fatigue assessment sits at the intersection of health technology, labor regulation, and data protection law. Many jurisdictions have begun to regulate high-risk artificial intelligence systems, requiring risk assessments, documentation, and human oversight. Fatigue assessment used in safety-critical employment contexts may fall within high-risk categories. The deployment of such systems should be accompanied by clear documentation of intended use, model limitations, training data characteristics, and validation procedures. Without such documentation, organizations may expose workers or users to unexplained decisions and may face legal liability. High-performance medicine and algorithmic decision-making have shown that integration into clinical and operational workflows requires more than prediction accuracy; it requires trust, transparency, and workflow alignment [16].

Interpretability is a key policy concern. In high-stakes settings, black-box models may be difficult to justify to users, regulators, and courts. Interpretable models can sometimes achieve comparable performance while allowing direct inspection of how features contribute to predictions [17]. However, interpretability alone does not ensure fairness or safety; an interpretable model can encode harmful correlations. A broader framework of responsible deployment includes model documentation, uncertainty quantification, stress testing, and incident reporting. The health care literature has articulated both the promise and the limitations of deep learning in clinical contexts, emphasizing that data quality and integration are often more significant than model architecture [18]. Fatigue assessment systems should adopt similar standards of evidence and validation before they are used to make consequential decisions.

Policy must also address sustainability and long-term maintenance. Machine learning models degrade over time as sensor hardware changes, user populations evolve, and environmental conditions shift. A monitoring system should include drift detection, periodic retraining, and post-market surveillance. Safe deployment of machine learning systems in health care requires continuous evaluation and updating protocols to address distribution shifts and rare events [19]. Without such operations, fatigue models may become unreliable or biased. Model updates themselves create governance challenges, because changes in performance and behavior must be communicated to affected users. The opacity of complex models can obscure these changes, compounding accountability problems [20]. Sustainable deployment therefore requires institutional capacity for model stewardship, not just initial algorithm development.

7. Conclusion

Machine learning-based fatigue assessment using continuous physiological and motion signals offers significant potential for improving safety, health, and performance in many domains. However, the development of trustworthy systems cannot be reduced to the optimization of prediction accuracy. The preceding analysis has emphasized that architectural choices, signal processing infrastructure, modeling trade-offs, deployment robustness, fairness,

and accountability are all integral to system performance. Fatigue assessment platforms operate within complex social and organizational settings, and their design must reflect these realities from the outset.

Future research should continue to develop sensing systems that are more energy-efficient and tolerant of real-world motion artifacts, while also creating standardized benchmarks that allow meaningful comparison across devices and populations. Privacy-preserving learning, causal modeling, and continual personalization are promising directions, but they must be accompanied by stronger governance mechanisms. Participatory design with workers, clinicians, regulators, and end users can help ensure that fatigue assessment technologies support human flourishing rather than merely extending surveillance. A mature field will require not only better machine learning models but also durable institutional frameworks for data stewardship, model auditing, and long-term operational integrity.

References

1. Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal sensing: Understanding mental health using ubiquitous sensors and machine learning. Annual Review of Clinical Psychology, 13, 23-47.

2. Banaee, H., Ahmed, M. U., & Loutfi, A. (2013). Data mining for wearable sensors in health monitoring systems: A review of recent trends and challenges. Sensors, 13(12), 17472-17500.

3. Reiss, A., & Stricker, D. (2012). Introducing a new benchmarked dataset for activity monitoring. In 2012 16th International Symposium on Wearable Computers (pp. 108-109). IEEE.

4. Mukhopadhyay, S. C. (2015). Wearable sensors for human activity monitoring: A review. IEEE Sensors Journal, 15(3), 1321-1330.

5. 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.

6. Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement, 28(3), R1-R39.

7. Schäfer, A., & Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. International Journal of Cardiology, 166(1), 15-29.

8. Healey, J. A., & Picard, R. W. (2005). Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on Intelligent Transportation Systems, 6(2), 156-166.

9. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM.

10. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

11. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information

Processing Systems 30 (pp. 5998-6008). Neural Information Processing Systems Foundation.

12. Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50-60.
13. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.
14. 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.
15. 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.
16. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.
17. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.
18. 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.
19. Saria, S., & Subbaswamy, A. (2019). Tutorial: Safe and reliable machine learning for healthcare. arXiv preprint arXiv:1902.07204.
20. Castelvecchi, D. (2016). Can we open the black box of AI? Nature, 538(7623), 20-23.