

# Robust Feature Learning for Stress and Fatigue Recognition from Noisy Wearable Physiological Signals

Anish J. Saxena

Department of Computer Science, Colorado State University, Fort Collins, CO, USA.  
saxena746@colostate.edu

Stefano Weber

School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.  
weber431@oregonstate.edu

## Abstract

Wearable physiological sensing has become a central component of modern health monitoring, offering continuous access to signals that are relevant for stress and fatigue recognition. However, the practical utility of these systems is constrained by motion artifacts, sensor drift, environmental interference, and variability in skin contact, all of which introduce complex noise patterns that can undermine learned representations. This paper examines robust feature learning for stress and fatigue recognition from noisy wearable physiological signals from a system-level perspective. Rather than focusing narrowly on algorithmic performance, the discussion integrates architectural decisions, data governance, deployment constraints, sustainability, fairness, and policy implications. The paper analyzes the structural trade-offs associated with shallow and deep feature learners, temporal modeling, transfer learning, and domain adaptation under realistic sensing conditions. It further considers how data provenance, annotation uncertainty, demographic representation, and regulatory expectations shape the design of robust recognition infrastructures. The paper argues that robustness cannot be achieved solely through model refinement, but requires coordinated attention to sensing hardware, signal preprocessing, feature representation, validation protocols, and organizational oversight. A cross-domain perspective is adopted, drawing on experiences in clinical monitoring, human factors engineering, and large-scale artificial intelligence systems. The conclusion emphasizes that sustainable stress and fatigue recognition systems must be evaluated not only by aggregate accuracy, but also by their behavior under distributional shift, their resource demands, their interpretability, and their alignment with social and ethical standards.

## Keywords

stress recognition, fatigue recognition, wearable physiological signals, robust feature learning, noise, system governance, data fairness, deployment sustainability.

## 1. Introduction

The growing availability of wearable devices has transformed physiological monitoring from episodic clinical assessments into continuous, everyday practice. Wrist-worn, chest-mounted, and head-worn sensors now collect photoplethysmographic signals, electrodermal activity, electrocardiographic traces, skin temperature, and movement data at scale. These

developments create opportunities for recognizing stress and fatigue in real time, with potential applications in occupational health, transportation safety, clinical care, and personal well-being. However, the transition from controlled laboratory protocols to free-living conditions introduces substantial uncertainty. Deep learning and related representation learning methods have shown considerable promise in extracting patterns from raw and minimally processed physiological data [1, 2]. Healthcare-oriented machine learning systems have also demonstrated that learned representations can capture subtle physiological states when trained on sufficiently diverse datasets [3].

Despite these advances, robust recognition of stress and fatigue remains difficult because wearable signals are inherently noisy. Motion artifacts, varying sensor contact, environmental light, sweating, and electromagnetic interference combine to produce complex and nonstationary disturbances. A model that performs well on clean laboratory data may fail when deployed in an occupational setting where users move continuously or in a clinical setting where sensors are attached by different personnel. Robust feature learning must therefore be understood as a systems problem rather than a narrow modeling problem. The architecture of the learning system, the preprocessing pipeline, the data governance framework, the evaluation strategy, and the deployment infrastructure all interact to determine whether a stress or fatigue recognition system can maintain acceptable performance over time.

This paper presents a long-form system-level analysis of robust feature learning for stress and fatigue recognition from noisy wearable physiological signals. It examines the nature of noise in wearable sensing, discusses representative feature learning architectures, and explores the structural trade-offs involved in designing systems that are both accurate and dependable. The paper also addresses fairness, interpretability, sustainability, and policy dimensions, because these factors increasingly determine whether such systems can be responsibly integrated into real-world institutions. The aim is not to propose a single algorithmic solution, but to provide an interdisciplinary perspective that connects signal processing, machine learning, infrastructure design, and governance.

## **2. Problem Context and System-Level Challenges**

Stress and fatigue recognition from wearable signals is not simply a classification task with fixed input-output pairs. It is a longitudinal inference problem embedded in a complex socio-technical environment. The physiological manifestations of stress and fatigue vary across individuals, contexts, and timescales. Acute stress may appear as rapid changes in electrodermal activity, while chronic fatigue may be reflected in altered heart rate variability over longer periods. The same physiological pattern can arise from physical exertion, emotional activation, illness, or environmental conditions. This ambiguity imposes strong demands on feature learning systems, which must extract representations that are both discriminative and robust to confounding factors.

From a system-level perspective, the selection of a learning approach involves trade-offs among interpretability, computational cost, sensitivity to noise, and adaptability. Traditional feature engineering with tree-based ensembles such as gradient boosting can provide strong performance on tabular physiological summaries and is often computationally efficient [4]. However, such models may depend heavily on handcrafted features that are sensitive to artifact contamination. Recurrent neural models, including long short-term memory architectures, can model temporal dependencies in physiological sequences and have been widely used for state recognition [5]. Attention-based models further allow the system to weight informative segments of a signal, which is valuable when portions of the recording are

corrupted or irrelevant [6]. Yet these deeper models require more data, more careful regularization, and more substantial computational resources.

The system-level challenge is thus to balance model capacity with the realities of deployment. A highly expressive model may learn spurious noise patterns if the training data contain systematic artifacts. Conversely, a simpler model may fail to capture the nuanced temporal dynamics of stress and fatigue. Robustness emerges from a combination of source characterization, preprocessing, representation learning, and evaluation under realistic perturbations. This requires shared infrastructure that can measure noise conditions, annotate data quality, and monitor model behavior over time. Without such infrastructure, model improvements may be fragile and short-lived.

### **3. Noise and Artifact Landscape in Wearable Physiological Sensing**

Wearable physiological signals are corrupted by multiple interacting noise sources. Motion artifacts are particularly problematic because physical movement can cause large transient distortions in photoplethysmographic and electrodermal recordings. These artifacts often overlap in frequency with the physiological features of interest, making simple filtering insufficient. Sensor displacement, changes in contact pressure, and skin stretch can alter the amplitude and baseline of signals. Environmental factors, including temperature and ambient light, introduce slow drifts and occasional outliers. In addition, device-level differences in sampling rate, gain, and filtering create distributional shifts across hardware platforms.

Robust feature learning must therefore operate in a setting where the boundary between signal and noise is ambiguous. A major structural decision is whether to perform explicit artifact detection and removal before representation learning or to learn representations that are intrinsically tolerant to artifact patterns. The first approach emphasizes preprocessing transparency and can incorporate domain knowledge about physiological signal characteristics. The second approach allows the model to adapt to complex artifact distributions but may obscure the distinction between meaningful physiological variation and noise. In practice, hybrid systems are often most effective. For example, contextual representations developed in related sequence modeling fields can be transferred to wearable data, while residual learning strategies can help deep networks maintain stable gradients in the presence of corrupted input segments [7, 8]. Transfer from medical imaging has also shown that representations learned in one high-dimensional domain can be adapted to related biomedical tasks [9].

Adaptive filtering techniques are an important part of this landscape. Adaptive notch-filtration methods have been shown to recover photoplethysmographic signals during physical activity by suppressing artifact-dominated frequency components while preserving physiological information [10]. Such methods illustrate a broader principle: robustness depends on the ability to adjust processing dynamically in response to changing signal characteristics. Rather than treating noise as a static distortion, the system must model noise as a context-dependent process. Feature learners that are coupled with such adaptive preprocessing can benefit from cleaner input representations, but they must also be trained to handle residual artifacts that survive preprocessing.

### **4. Feature Learning Architectures for Robust Representation**

Feature learning for physiological time series can be organized along several architectural dimensions: the depth of representation, the temporal receptive field, the degree of supervision, and the use of domain adaptation. Shallow representations based on statistical

summaries and frequency-domain features provide interpretability and lower computational cost, but they are often brittle under nonstationary noise. Deep convolutional and recurrent architectures can learn hierarchical features that separate slow physiological trends from high-frequency artifacts, provided that training data cover diverse noise conditions. However, deep models can also overfit to device-specific noise characteristics, leading to poor generalization when sensors or populations change.

Transfer learning offers a principled way to reuse representations across related tasks and domains [11]. In stress and fatigue recognition, a model pretrained on large-scale physiological data can be fine-tuned for a specific application with limited labeled examples. This strategy reduces the need for costly data collection and can improve robustness by exposing the model to a broader range of signal variability. Domain adaptation further addresses the mismatch between training and deployment conditions by aligning feature distributions across source and target domains [12]. Adversarial training approaches can encourage the feature extractor to produce representations that are invariant to noise conditions, sensor type, or user activity, while preserving information needed for state recognition.

Architectural decisions also interact with deployment constraints. Lightweight models may be necessary for on-device inference to reduce latency and preserve privacy. However, aggressive compression can reduce the capacity needed to handle complex artifacts. System designers must consider whether inference will occur on the wearable device, on a paired smartphone, or in a cloud environment. Each option has implications for energy consumption, data transmission, privacy, and update frequency. A cloud-based system can host larger models and enable continuous learning, but it raises concerns about connectivity and data governance. On-device inference offers lower latency and stronger privacy, but limits model complexity and complicates centralized monitoring. These trade-offs are central to the design of sustainable and robust wearable recognition systems.

## **5. Data Governance, Fairness, and Deployment Considerations**

Robustness in stress and fatigue recognition is not solely a technical property; it is also shaped by data governance and fairness. Wearable physiological data are deeply personal, and their collection, storage, and analysis raise significant privacy concerns. The expansion of continuous health monitoring intersects with broader debates about surveillance and data commodification [13]. Institutions deploying these systems must establish clear policies for consent, data minimization, retention, and access. Governance frameworks should also address secondary uses of physiological data, including employment decisions, insurance underwriting, and behavioral profiling.

Fairness is another critical dimension. Stress and fatigue recognition models can perform differently across demographic groups due to variations in skin tone, skin conductance, baseline physiology, and device placement. If training data are dominated by a narrow population, the learned representations may not generalize equitably. A unified framework for ethical artificial intelligence emphasizes beneficence, non-maleficence, autonomy, justice, and explicability [14]. Algorithmic systems that make consequential inferences about individuals must be audited for differential performance and unintended harms [15]. In clinical and occupational settings, inaccurate stress or fatigue predictions can affect patient safety, workload allocation, and professional judgment. Machine learning tools in medicine can provide substantial value, but they require careful validation and contextual integration [16].

The deployment of stress and fatigue recognition systems should therefore include fairness audits, demographic performance reporting, and clear procedures for contesting automated inferences. High-performance medical artificial intelligence systems increasingly rely on multimodal data and deep representations [17]. Yet the translation of these systems into practice requires more than predictive accuracy. It requires confidence in system behavior under real-world variability, trust among users and clinicians, and alignment with institutional accountability structures.

## **6. Infrastructure and Sustainability Dimensions**

Sustainable deployment of robust feature learning systems depends on underlying infrastructure. Data pipelines must handle heterogeneous devices, variable sampling rates, missing data, and irregular upload patterns. Labeling stress and fatigue often relies on subjective self-reports, ecological momentary assessments, or expert annotations, all of which introduce uncertainty. The infrastructure should support data quality tracking, provenance recording, and model versioning so that performance can be reproduced and audited. Modular architectures that separate data ingestion, preprocessing, feature learning, inference, and monitoring can improve maintainability and allow components to be updated independently.

Computational sustainability is increasingly important. Training large deep models on long physiological recordings consumes significant energy and may be impractical for many organizations. Model distillation, parameter sharing, and selective fine-tuning can reduce resource demands. At the same time, robust systems may require online adaptation, which introduces additional computational costs. Designers should evaluate whether performance gains justify the environmental and operational costs of continuous retraining. Lightweight feature extractors that can run efficiently on embedded hardware may be more sustainable than cloud-dependent models, particularly in low-resource settings.

Interoperability is another system-level concern. Wearable data formats and application programming interfaces vary widely. Standardization efforts can reduce integration overhead and enable cross-institutional validation. Without interoperable data models, robust feature learning systems may become locked into proprietary ecosystems, limiting their generalizability and societal benefit. The governance of these infrastructures should promote open benchmarks, shared data quality metrics, and transparent reporting standards that allow meaningful comparisons across studies and devices.

## **7. Evaluation, Validation, and Cross-Domain Transfer**

Evaluation of stress and fatigue recognition systems must go beyond aggregate accuracy. Because noise conditions vary across environments, evaluation should include multiple perturbation scenarios such as simulated motion artifacts, sensor dropout, and cross-device transfer. Systems should be assessed on their ability to maintain calibration under distributional shift, not merely on overall classification rates. Interpretability methods can help identify whether a model relies on physiologically plausible features or on spurious noise patterns. Surveys of interpretability techniques highlight the diversity of approaches available for inspecting deep representations [18]. Visual explanation methods can localize salient segments in time series or highlight regions of input that drive predictions [19]. Additive feature attribution methods provide a unified framework for explaining individual predictions and can support error analysis in clinical contexts [20].

Cross-domain transfer is particularly important for stress and fatigue recognition. A system trained in a laboratory environment may fail when deployed in an occupational setting with

different movement patterns. Dense connectivity in deep convolutional architectures can improve gradient flow and feature reuse, which may support more stable learning from limited physiological data [21]. Axiomatic attribution techniques can help identify whether transferred models retain meaningful physiological structure after fine-tuning [22]. Systematic evaluation should include multiple target populations, devices, and contexts to detect hidden fragility. It should also include longitudinal assessment, because stress and fatigue patterns evolve over time and may be influenced by seasonal, occupational, or health-related shifts.

A broader perspective from algorithmic fairness research emphasizes that robustness and fairness are interconnected. Models that are brittle under distributional shift often exhibit differential performance across groups [23]. Therefore, robustness audits should be coupled with fairness audits. If a model fails more often for users with darker skin tones because photoplethysmographic signal quality differs, this is both a technical and an ethical failure. Evaluation protocols should explicitly measure such disparities and require remediation before deployment.

## **8. Policy Implications and Future Directions**

Policy frameworks for wearable-based stress and fatigue recognition are still evolving. Regulatory bodies increasingly expect evidence of safety, effectiveness, and equitable performance for software-based medical and wellness interventions. A robust feature learning system should be accompanied by documentation that describes intended use, target populations, data collection procedures, model limitations, and monitoring plans. This documentation supports institutional accountability and enables independent review. Policy makers should also consider the implications of using fatigue recognition in safety-critical industries such as transportation, construction, and healthcare. While early warning systems may reduce accidents, they may also create new forms of surveillance and worker monitoring that require careful legal and ethical scrutiny.

Future research should develop standardized noise benchmarks that reflect the complexity of free-living conditions. These benchmarks would allow researchers to compare feature learning methods under realistic artifact profiles and promote cumulative progress. Federated learning and privacy-preserving analytics can enable multi-institution collaboration without centralizing sensitive physiological data. However, federated systems introduce new challenges in heterogeneity, communication efficiency, and secure aggregation. System-level research should address these challenges alongside model architecture design.

Another direction is the integration of physiological sensing with contextual information such as work schedules, sleep history, and environmental conditions. Contextual data can reduce ambiguity in stress and fatigue recognition, but they also raise privacy concerns and may introduce new biases. The design of such multimodal systems must balance predictive value with data minimization and user autonomy. Ultimately, robust feature learning will be achieved not through a single architectural breakthrough, but through the coordinated evolution of sensing hardware, data infrastructure, learning algorithms, governance practices, and policy standards.

## **9. Conclusion**

Robust feature learning for stress and fatigue recognition from noisy wearable physiological signals is a multifaceted challenge that extends well beyond algorithm design. Wearable signals are contaminated by motion artifacts, sensor drift, and environmental interference that vary across individuals and deployment contexts. Feature learning architectures must

therefore be designed to handle nonstationary noise while preserving physiologically meaningful patterns. Transfer learning, domain adaptation, adaptive filtering, and lightweight deep architectures offer complementary strategies for improving robustness. However, these technical approaches must be embedded within a broader infrastructure that supports data quality tracking, fairness auditing, interpretability, and sustainable deployment.

This paper has argued that system-level considerations are inseparable from model-level performance. Data governance frameworks determine how physiological data are collected and used. Fairness audits reveal whether recognition systems perform equitably across populations. Deployment choices shape energy consumption, latency, privacy, and maintainability. Evaluation protocols must include realistic noise perturbations and cross-domain transfer scenarios. Policy mechanisms must ensure that stress and fatigue recognition technologies are deployed with appropriate oversight and accountability.

The future of wearable stress and fatigue recognition lies not in treating noise as a nuisance to be eliminated, but as an intrinsic characteristic of real-world sensing that must be structurally accommodated. This requires interdisciplinary collaboration across signal processing, machine learning, human factors, ethics, and infrastructure engineering. By adopting a system-level perspective, researchers and practitioners can build recognition systems that are not only accurate in controlled settings but also dependable, fair, and sustainable in the complex environments where they are ultimately needed.

## References

1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.
2. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
3. 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.
4. 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.
5. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
7. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
8. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
9. Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems*, 32, 3347–3357.

10. Zheng, X., Dwyer, V. M., Barrett, L. A., Derakhshani, M., & Hu, S. (2022). Adaptive notch-filtration to effectively recover photoplethysmographic signals during physical activity. *Biomedical Signal Processing and Control*, 72, 103303.
11. Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.
12. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(1), 2096–2030.
13. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. *PublicAffairs*.
14. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1).
15. Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21.
16. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358.
17. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44–56.
18. Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., Srivastava, M., Preece, A., Julier, S., Rao, R. M., Kelley, T. D., Braines, D., Sensoy, M., Willis, C. J., & Gurram, P. (2017). Interpretability of deep learning models: A survey of results. *Proceedings of IEEE SmartWorld*, 1–6.
19. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*, 618–626.
20. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
21. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.
22. Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. *Proceedings of the 34th International Conference on Machine Learning*, 3319–3328.
23. 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.