

# Self-Supervised Representation Learning for Motion-Resilient Wearable Biosignal Analysis

Langcong Gao

School of Electrical Engineering and Computer Science, Oregon State University, Corvallis,  
OR, USA.

langcong2003@oregonstate.edu

Fernando J. Virtanen

School of Computing, Clemson University, Clemson, SC, USA.

fernandowork@clemson.edu

## Abstract

Wearable biosignal analysis has become a critical infrastructure for continuous health monitoring, yet motion artifacts, heterogeneous sensor configurations, and scarce clinical labels continue to limit the generalization of supervised deep models. This paper examines self-supervised representation learning as a system-level strategy for building motion-resilient biosignal analysis pipelines. Rather than focusing narrowly on model architectures, the discussion addresses the structural trade-offs among on-body sensing, edge computing, cloud aggregation, representation pretraining, downstream adaptation, and governance mechanisms. The central argument is that motion resilience is not solely a signal processing problem but an emergent property of a learning system that combines contrastive pretraining, temporal augmentation, multimodal synchronization, and context-aware deployment. The paper analyzes how representations learned from unlabeled ambulatory data can reduce dependence on curated datasets, support transfer across sensor placements and activity contexts, and improve robustness under distributional shift. It further considers infrastructure requirements, computational sustainability, regulatory constraints, fairness risks, and privacy-preserving federated learning. Drawing on recent advances in contrastive learning, physiological signal modeling, and decentralized systems, the paper provides a forward-looking perspective on the design of scalable, equitable, and governable biosignal learning infrastructures. The discussion emphasizes that technical performance must be evaluated together with deployment feasibility, energy efficiency, data governance, and societal implications to ensure that self-supervised biosignal representations are safe and effective outside controlled laboratory settings.

## Keywords

wearable biosignals; self-supervised learning; motion artifacts; representation learning; federated governance; fairness.

## 1. Introduction

The proliferation of wearable devices has transformed biosignal monitoring from episodic clinical measurements into continuous ambulatory observation across daily life. Electrocardiography, photoplethysmography, electrodermal activity, and electromyography are now collected in unstructured environments where movement, sensor displacement, environmental illumination, and varying skin contact introduce severe artifacts. Deep learning has demonstrated remarkable capacity to model complex physiological patterns [1]. However,

the benefits of deep models depend on representation quality and the alignment between training distributions and deployment conditions. Representation learning has therefore become a central research direction because it seeks to capture the underlying factors of variation in biosignal data in a form that can be reused across tasks, users, and sensor configurations [2]. The challenge is that biosignal data in ambulatory settings are abundant but largely unlabeled, while clinical annotations remain expensive, sparse, and often tied to controlled protocols.

Conventional supervised learning pipelines have relied on deep architectures such as residual networks to extract hierarchical features from physiological time series [3]. More recently, self-supervised pretraining has transformed natural language processing by learning general-purpose representations from large unlabeled corpora [4]. In computer vision, contrastive learning frameworks have shown that representations learned by maximizing agreement between augmented views can rival supervised representations on downstream tasks [5]. These advances are highly relevant to wearable biosignal analysis, where large-scale public physiological repositories provide varied recordings but only partially labeled records [6]. In parallel, attention-based mechanisms have been proposed to improve physiological monitoring during physical activity by jointly modeling multiple signal channels and motion-related contextual cues [7]. Such approaches indicate the potential of architectures that explicitly account for motion-induced interference, but deploying them at scale requires a broader systems perspective.

A complementary line of work has formulated self-supervised representation learning through contrastive predictive coding, which learns to predict future latent states from past observations [8]. For wearable biosignals, predictive objectives can exploit temporal structure, cross-sensor redundancy, and subject-specific baselines to learn representations that are less sensitive to transient artifacts. This paper examines motion-resilient wearable biosignal analysis as a socio-technical infrastructure problem. It considers how representation learning interacts with sensing hardware, communication constraints, edge processing, cloud orchestration, model governance, and equity. The core argument is that motion resilience should be designed as an emergent property of the entire system rather than isolated filtering or artifact removal. The remaining sections analyze background research, system architecture, representation design, robustness, deployment governance, and policy implications.

## **2. Background and Related Work**

Self-supervised methods for healthcare have developed rapidly because annotations are difficult to obtain and often require expert clinical review. Contrastive learning in human activity recognition has been explored as a way to leverage unlabeled wearable sensor streams to learn representations that transfer to downstream tasks such as gesture recognition and health monitoring [9]. For cardiac signals, contrastive objectives across spatial leads, temporal windows, and patients have been used to learn representations that improve classification of arrhythmias with limited labeled data [10]. Similar methods applied to twelve-lead electrocardiogram data have shown that self-supervised pretraining can outperform supervised baselines when fine-tuned on small clinical datasets [11]. A broader review of self-supervised learning in medicine highlights both the promise and the need for careful validation because medical data exhibit strong correlations, hidden confounders, and distributional shifts that may be exploited by proxy objectives [12]. These findings provide the foundation for motion-resilient biosignal representation, but most studies focus on diagnostic accuracy rather than deployment-level concerns.

Transfer learning has been extensively studied as a framework for adapting representations learned in one domain to related tasks in another [13]. In wearable biosignal systems, transfer is necessary because sensor placements, sampling rates, skin properties, and motion contexts vary widely across individuals and devices. Federated learning offers a decentralized training model that can leverage distributed data without centralizing raw physiological records [14]. In medical settings, federated approaches have been used to train models across institutions while preserving data locality and complying with data sharing restrictions [15]. These decentralized methods are particularly relevant to ambulatory biosignal analysis because continuous wearable data are sensitive and governed by multiple jurisdictional frameworks. However, federated learning introduces additional system complexity, including communication overhead, non-IID data distributions across clients, and device heterogeneity that demand careful system design [16]. Domain adaptation techniques that align feature representations across users, devices, and environmental contexts have accordingly assumed critical importance for transferring laboratory-derived models to ambulatory settings [17]. The system-level challenge is therefore not merely to learn invariant features but to coordinate these learning processes across heterogeneous sensing nodes, edge devices, and central aggregation points while preserving clinical validity.

### **3. System Architecture for Motion-Resilient Biosignal Analysis**

A motion-resilient biosignal analysis system can be understood as a layered infrastructure in which sensing, representation learning, adaptation, and decision-making are distributed across multiple physical and organizational boundaries. At the lowest layer, multimodal wearable sensors collect physiological signals such as electrocardiogram, photoplethysmography, electrodermal activity, and accelerometry. The simultaneous availability of motion-related channels provides an important self-supervisory signal because accelerometer readings can be used to construct augmentations or to identify periods of high motion artifact without requiring manual labels [18]. For example, wrist-worn photoplethysmography is particularly susceptible to hand movement, and combining optical signals with inertial measurements has been shown to improve heart rate estimation during intensive physical exercise [19]. This sensor fusion principle extends naturally to self-supervised representation learning, where the objective is not to reconstruct a single channel but to learn cross-modal correspondences that are robust to transient disturbances.

Above the sensing layer, the representation encoder transforms raw or lightly preprocessed signals into compact embeddings that can be reused across multiple downstream tasks. In a centralized design, all raw data would be transmitted to a cloud server for pretraining, but this model is often impractical because of bandwidth, latency, energy, and privacy constraints. Edge computing offers an alternative in which initial feature extraction and artifact suppression occur close to the body, while only aggregated embeddings or model updates are sent to more powerful servers [20]. The architectural trade-off is between local adaptivity and global consistency. Edge-local representations can be tuned to an individual user's gait, skin tone, sensor placement, and activity patterns, but they may overfit to idiosyncratic conditions and fail to generalize. Cloud-side aggregation, by contrast, enables cross-user statistical strength but risks homogenizing away clinically meaningful individual variation. A well-designed system therefore requires a hierarchical representation strategy in which low-level artifact-invariant features are learned locally, while higher-level semantic features are aligned across the population through federated or split learning mechanisms.

The governance of this distributed architecture depends on carefully specified interfaces among components. Each layer must expose a contract regarding data modality, sampling rate, embedding dimensionality, update frequency, and privacy budget. Without such contracts, motion resilience can degrade silently when sensors are repositioned, firmware is updated, or users change devices. The embedding space must also remain interpretable enough for clinicians and regulators to audit failure modes. This does not imply that all internal representations must be human-readable, but the system should support selective visualization, saliency analysis, and out-of-distribution detection at the boundaries where decisions are made. The structural design thus links technical robustness to accountability, because a representation that fails gracefully under motion artifact should also produce recognizable uncertainty signals rather than confident but erroneous predictions.

#### **4. Self-Supervised Pretraining and Augmentation Strategies**

The core learning mechanism for motion-resilient biosignal analysis is the design of self-supervised objectives that do not require manual annotations yet still encourage the model to discard nuisance variation introduced by movement. Contrastive learning, predictive coding, and masked reconstruction are three families of objectives that have shown particular promise. Contrastive methods encourage different augmented views of the same underlying physiological state to map to nearby embeddings, while views drawn from different temporal contexts or subjects are pushed apart [5]. In wearable biosignals, augmentations must be physiologically plausible. Random cropping, temporal warping, channel dropout, and noise injection can simulate moderate sensor displacement, but excessive transformation may destroy diagnostic information. Motion-specific augmentations can be generated by mixing clean recordings with synthetic motion artifacts derived from accelerometer records, thereby teaching the model to separate physiological signal from movement interference.

Predictive objectives offer a complementary signal by leveraging the temporal coherence of physiological processes. A representation that predicts future latent states from past observations must capture slow-moving cardiovascular or autonomic dynamics while ignoring high-frequency motion transients [8]. This approach is especially attractive for long-term ambulatory monitoring because it naturally exploits the sequential structure of wearable data. However, predictive objectives may inadvertently learn to rely on periodicity that is disrupted by arrhythmias or other pathological events, so they must be combined with contrastive or reconstruction losses that preserve local diagnostic detail. Masked reconstruction, in which portions of the input signal are hidden and the model must infer them from context, provides another route to learning robust features. When applied to multichannel biosignals, masking one modality and reconstructing it from others encourages cross-modal redundancy that can compensate for temporary sensor failure.

The pretraining dataset itself is a critical infrastructure component. Public physiological repositories such as PhysioNet provide diverse recordings, but their coverage of ambulatory motion contexts remains uneven [6]. Dataset curation must therefore account for demographic diversity, activity type, sensor placement, skin pigmentation, and comorbidities. A pretraining corpus dominated by healthy young adults exercising in controlled laboratory settings will not yield representations that transfer safely to older adults with cardiovascular disease or to individuals with darker skin tones whose photoplethysmography signals are attenuated by melanin. Self-supervised objectives can amplify existing biases because they optimize for statistical regularities in the available data without external clinical correction. Consequently,

dataset governance and representation learning must be treated as a single problem rather than separate technical and ethical concerns.

## **5. Robustness, Fairness, and Domain Transfer**

Motion resilience is closely related to robustness under distributional shift, but the two concepts are not identical. Robustness in the narrow sense refers to maintaining accuracy when the input distribution changes due to motion artifacts, sensor drift, or environmental noise. Fairness, by contrast, requires that performance remains equitable across subpopulations defined by demographic, physiological, or contextual attributes. A representation may be highly robust on average while still failing disproportionately for users with certain gait patterns, skin tones, or body compositions. The distinction matters because self-supervised pretraining can inadvertently learn shortcuts that are correlated with protected attributes. For example, if photoplethysmography signal quality is lower for individuals with darker skin, the model may learn to rely more heavily on accelerometer-derived motion features for those users, thereby reducing diagnostic accuracy for cardiovascular events. System designers must therefore evaluate motion-resilient representations using disaggregated metrics that report performance separately across relevant subgroups.

Domain transfer is a further source of structural tension. Representations learned from wrist-worn sensors may not transfer directly to chest-mounted patches or earbud-based photoplethysmography because the underlying physiological signals and motion coupling differ substantially. Transfer learning research has shown that feature reuse is most effective when the source and target domains share low-level statistical regularities [13]. In biosignal analysis, this suggests that pretraining should include multiple anatomical sites and sensor modalities to encourage learning of body-wide physiological dynamics rather than site-specific artifacts. Domain-adversarial training can further align representations across sensor placements and activity contexts, but it must be applied with care to avoid removing clinically meaningful differences between individuals or disease states [17]. The optimal representation is not one that erases all domain information but one that separates nuisance variation from clinically relevant variation.

Fairness audits in biosignal systems face the additional challenge that ground truth labels may themselves be biased. Clinician annotations are often derived from populations that are not representative of the full deployment population, and manual review of ambulatory recordings may underdiagnose certain conditions in women or minority groups. Self-supervised learning does not require labels during pretraining, but downstream fine-tuning and evaluation still depend on annotated datasets. A motion-resilient system that improves performance for the majority while degrading performance for historically underserved groups would not be acceptable in a health monitoring context. Fairness must therefore be designed into the representation space through careful subsampling, adversarial debiasing, and transparent performance reporting. The governance mechanisms discussed later provide the institutional support for such practices.

## **6. Deployment, Edge Computing, and Sustainability**

The deployment of motion-resilient biosignal representations must contend with the energy and memory constraints of wearable devices. Deep transformer encoders and large contrastive backbones may achieve high accuracy in server-side evaluations but cannot be executed continuously on a wristband with a limited battery. This creates pressure to distill large pretrained models into compact architectures that retain artifact-invariant properties while

operating within strict computational budgets. Edge computing has been proposed as a way to offload intensive computation to nearby gateways such as smartphones or home hubs, but this approach introduces dependencies on network availability and device pairing [20]. A more sustainable design may combine a lightweight on-device encoder that performs real-time denoising with periodic cloud synchronization for representation updates and diagnostic model refresh.

Energy consumption is not only an engineering constraint but also an environmental and equity concern. Training large self-supervised models on physiological data can require substantial computational resources, and the carbon footprint of such training runs has drawn increasing scrutiny [21]. In the context of wearable health monitoring, the cumulative energy cost of billions of devices continuously transmitting data for cloud-based inference would be considerable. System designers should therefore adopt principles of green artificial intelligence, including efficient architectures, selective data transmission, and federated training schedules that reduce communication overhead [25]. These choices also have implications for global equity, because users in regions with unreliable electricity or expensive connectivity may be excluded from cloud-dependent services. A motion-resilient system that relies too heavily on centralized inference may thus reproduce or deepen existing health disparities.

The trade-off between local and remote computation is not purely technical but reflects governance decisions about where intelligence should reside. Edge-local models can provide lower latency and greater privacy, but they are harder to update, audit, and manage across large populations. Cloud-centralized models offer easier monitoring and consistent updates, but they create privacy risks and single points of failure. Hybrid architectures that distribute representation learning across layers must therefore specify clear protocols for model versioning, fallback behavior during connectivity loss, and differential privacy guarantees. Differential privacy mechanisms can be applied to federated updates to limit the leakage of individual physiological patterns while still allowing population-level learning [22]. However, the addition of noise to model updates may interact with motion-resilience objectives, because high-frequency artifacts can be confused with privacy-preserving perturbations. The system must carefully balance the perturbation required for privacy against the perturbation that motion artifacts already introduce.

## **7. Governance, Privacy, and Policy Implications**

The governance of self-supervised biosignal representation systems extends beyond technical validation to include regulatory compliance, institutional oversight, and public accountability. In many jurisdictions, software that provides diagnostic or therapeutic recommendations from physiological signals is classified as a medical device and subject to premarket review or postmarket surveillance. The U.S. Food and Drug Administration has articulated a regulatory framework for adaptive artificial intelligence systems, emphasizing the need for predetermined change control plans and real-world performance monitoring [24]. Self-supervised representation learning complicates this framework because the pretraining objective may evolve continuously as new unlabeled data are incorporated, even before any labeled downstream task changes. Regulators and developers must therefore establish clear boundaries between representation updates, which may occur through unsupervised learning, and clinical decision functions, which should require more rigorous validation and locking procedures.

Privacy is a central governance concern because ambulatory biosignals can reveal sensitive information about cardiac health, stress, sleep, substance use, and even emotional states. Federated learning has been widely proposed as a privacy-preserving alternative to centralized training [14]. However, federated systems are not automatically private; model updates can leak information about individual training examples through gradient inversion or membership inference attacks. Differential privacy provides a formal framework for limiting such leakage, but its rigorous application to time-series biosignals remains an open research problem [22]. Homomorphic encryption and secure aggregation can further reduce exposure during the aggregation step, but they add computational overhead that may be impractical on low-power wearables. Governance mechanisms must therefore define not only what data are collected but also what information is inferable from the model itself.

Algorithmic accountability is also essential for motion-resilient systems because failures often occur under challenging real-world conditions that are underrepresented in test sets. Model cards and structured reporting frameworks can document intended use, performance across subgroups, known limitations, and failure modes [23]. For biosignal representations, such documentation should include the distribution of motion contexts, sensor types, skin tones, and clinical conditions present in the pretraining and fine-tuning datasets. Without this transparency, clinicians and end users cannot assess whether a wearable device is likely to perform acceptably for their specific circumstances. In addition, incident reporting systems should be established so that users can flag episodes of inaccurate heart rate tracking or missed arrhythmia detection, enabling continuous postmarket learning without requiring centralized access to raw signals.

## **8. Conclusion**

Self-supervised representation learning offers a promising path toward motion-resilient wearable biosignal analysis by exploiting the vast amounts of unlabeled ambulatory data generated by modern sensors. However, the transition from benchmark accuracy to dependable real-world performance requires attention to system architecture, data governance, fairness, and sustainability. Contrastive, predictive, and masked objectives can learn representations that separate physiological signals from motion artifacts, but only if the pretraining corpus reflects the diversity of deployment conditions. Distributed architectures can preserve privacy and reduce communication costs, but they introduce new challenges related to non-IID data, energy constraints, and regulatory oversight. Robustness must be evaluated not only in aggregate but also across demographic and contextual subgroups, because motion artifacts and sensor failures do not affect all users equally. Ultimately, motion resilience is best understood as an emergent property of a socio-technical system rather than a property of a single model. Future research should therefore integrate representation learning with edge computing, federated governance, differential privacy, and sustainability principles to build biosignal infrastructures that are technically effective, socially equitable, and institutionally accountable.

## **References**

1. Hannun, A. Y., Rajpurkar, P., Haghpanahi, M., Tison, G. H., Bourn, C., Turakhia, M. P., & Ng, A. Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine*, 25(1), 65–69.

2. Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), 1798–1828.
3. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).
4. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171–4186).
5. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning* (pp. 1597–1607).
6. Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C.-K., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215–e220.
7. 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.
8. van den Oord, A., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.
9. Saeed, A., Ozcelebi, T., & Lukkien, J. (2019). Multi-task self-supervised learning for human activity detection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2), Article 61, 1–18.
10. Kiyasseh, D., Zhu, T., & Clifton, D. A. (2021). CLOCS: Contrastive learning of cardiac signals across space, time, and patients. In *Proceedings of the 38th International Conference on Machine Learning* (pp. 5606–5615).
11. Mehari, T., & Strodthoff, N. (2022). Self-supervised representation learning from 12-lead ECG data. *Computers in Biology and Medicine*, 141, 105114.
12. Krishnan, R., Rajpurkar, P., & Topol, E. J. (2022). Self-supervised learning in medicine and healthcare. *Nature Biomedical Engineering*, 6(12), 1346–1352.
13. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1), 43–76.
14. McMahan, B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (pp. 1273–1282).
15. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. *NPJ Digital Medicine*, 3, Article 119.

16. 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.
17. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17(59), 1–35.
18. Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors*, 16(1), 115.
19. Zhang, Z., Pi, Z., & Liu, B. (2015). TROIKA: A general framework for heart rate monitoring using wrist-type photoplethysmographic signals during intensive physical exercise. *IEEE Transactions on Biomedical Engineering*, 62(2), 522–531.
20. Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). Edge computing: Vision and challenges. *IEEE Internet of Things Journal*, 3(5), 637–646.
21. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3645–3650).
22. Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proceedings of the Third Theory of Cryptography Conference* (pp. 265–284).
23. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 220–229).
24. U.S. Food and Drug Administration. (2021). Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan.
25. Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, 63(12), 54–63.