Multi-Label Health State Retrieval from Wearable Biosignals Through Statistical-Prior Guided Representation Learning and Deep Hash Encoding

Logan Sanders
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
logan.work@buffalo.edu

Malcolm C. Kelley
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
malcolmk@uab.edu

Abstract

The proliferation of wearable devices capable of continuously capturing photoplethysmography, electrocardiography, and other biosignals has created an urgent need for systems that can retrieve comprehensive health states from exponentially growing data streams. This paper presents a system-level investigation of multi-label health state retrieval that couples statistical-prior guided representation learning with deep hash encoding to achieve efficient and clinically meaningful search across large-scale biosignal repositories. The architecture is designed around a dual objective: first, a foundation modeling backbone integrates physiological and distributional priors to learn robust, generalizable signal representations that remain invariant to sensor noise, inter-individual variability, and temporal dynamics; second, a deep hashing module maps these representations into compact binary codes that support sublinear retrieval while preserving fine-grained semantic relationships, including comorbidity patterns and simultaneous physiological conditions. We discuss the structural trade-offs inherent in balancing code length, retrieval recall, and energy consumption, and analyze infrastructure choices spanning on-device inference, edge-cloud coordination, and federated learning for privacy-preserving model updates. Considerable attention is devoted to fairness, bias amplification across demographic and skin-tone groups, and the governance implications of deploying retrieval-driven decision support in regulated healthcare ecosystems. Through cross-domain comparisons with image and text retrieval, we highlight unique challenges such as physiological concept drift, multi-label granularity, and the interpretability of hash-based similarity. The discussion culminates in a forward-looking perspective on sustainable deployment, clinical validation pathways, and the alignment of multi-label retrieval systems with evolving regulatory frameworks.

Keywords

wearable biosignals, multi-label retrieval, representation learning, deep hashing, statistical priors, fairness, health informatics.

1. Introduction

Wearable sensing has transitioned from niche fitness tracking to a cornerstone of remote patient monitoring, preventive medicine, and large-scale epidemiological cohort studies. The resulting biosignal streams, predominantly photoplethysmography and electrocardiography, encode rich information about cardiovascular dynamics, respiratory patterns, autonomic regulation, and early signs of decompensation. Unlike traditional diagnostic pipelines that reduce each recording to a single disease label, contemporary clinical understanding emphasizes the multiplicity and co-occurrence of health states: an individual may simultaneously present with atrial fibrillation, sleep apnea, and hypertension, each leaving distinct but intertwined signatures in the biosignal morphology. Consequently, the ability to retrieve all relevant health states from a short wearable recording, rather than classifying it into a single category, has profound implications for decision support, retrospective analysis, and population health management. Multi-label health state retrieval from biosignals thus emerges as a systems challenge that intersects representation learning, efficient similarity search, and responsible deployment.

The core difficulty lies in simultaneously addressing semantic granularity and retrieval scalability. Wearable biosignals exhibit enormous inter-subject variability due to differences in skin pigmentation, sensor placement, motion artifacts, and underlying physiology. Learning representations that are sensitive enough to resolve subtle multi-label structure while being invariant to nuisance factors demands inductive biases that go beyond generic data augmentation. Statistical priors derived from physiological models, hemodynamic principles, and population-level distributions offer a principled route to embedding such inductive biases into foundation model training. When combined with deep hash encoding, which compresses high-dimensional representations into binary codes that can be compared in Hamming space with constant-time complexity, the resulting system enables rapid retrieval of health states from databases containing millions of recordings, even on resource-constrained edge devices.

This paper provides a system-oriented examination of this synergetic approach. We analyze the architectural principles, structural trade-offs, and socio-technical dimensions of building multi-label health state retrieval pipelines that unite statistical-prior guided representation learning with deep hashing. The discussion is organized around the entire lifecycle of such a system: acquisition and preprocessing, representation pretraining, hash function design, retrieval infrastructure, bias mitigation, and governance. We avoid mathematical formalisms in favor of conceptual depth, drawing on cross-domain lessons from information retrieval, biomedical engineering, and fairness-aware machine learning. Through this lens, we articulate how these components must be co-designed to satisfy the stringent requirements of clinical trustworthiness, energy efficiency, and regulatory compliance.

2. Background and System Motivations

Current wearable analytics systems predominantly operate in a classification paradigm, where a deep neural network outputs a probability distribution over a predefined set of diseases or wellness states. While this suffices for single-label screening, it fails to capture comorbid conditions without combinatorial explosion of output classes. Multi-label retrieval reframes the problem: a query signal is mapped into an embedding space, and the system searches a database of indexed recordings whose associated health labels are known, returning both the retrieved neighbors and the aggregated label sets. This approach treats labels as a set-valued target, naturally accommodating the co-occurrence structure of physiological dysregulation. The retrieval perspective also enables case-based reasoning, allowing clinicians to inspect similar historical cases together with their outcomes, thereby improving transparency.

Representation learning for biosignals has recently been revolutionized by self-supervised pretraining and foundation model paradigms adapted from natural language processing and computer vision. Contrastive methods that maximize agreement between differently augmented views of the same recording have yielded representations that transfer effectively to downstream arrhythmia classification and sleep staging. Yet such generic augmentations, such as random cropping or additive Gaussian noise, may violate physiological plausibility. A growing body of work therefore advocates for augmentations informed by signal generation models, such as pulse wave reflection theory or respiratory sinus arrhythmia dynamics. By incorporating statistical priors about waveform morphology, heart rate variability distributions, and noise characteristics, the learned representations acquire robustness to covariate shifts that plague real-world deployments, from changes in ambient temperature to sensor manufacturing variations.

Efficient retrieval in high-dimensional spaces has been extensively studied in multimedia contexts, where learning to hash has matured as a paradigm that jointly optimizes representation quality and binary code discriminability. Deep hashing networks project input examples into low-dimensional Hamming space such that semantic similarity is preserved, while dissimilar items are pushed apart. Key innovations include asymmetric pairwise constraints, margin-adaptive quantization, and self-supervised semantic excavation that can operate without exhaustive pairwise label annotations. Translating these techniques to biosignals introduces unique requirements: hash codes must encode fine-grained comorbidity relationships, accommodate multi-label similarity metrics such as Jaccard distance or ranking-based losses, and remain stable under temporal concept drift as a person's health state evolves.

The system motivation for combining statistical priors with deep hashing is twofold. On the one hand, priors improve representation quality under scarce labeled data and domain shift, directly benefiting the semantic fidelity of the resulting hash codes. On the other hand, deep hashing provides the retrieval efficiency needed to deploy such models on wearables or edge servers with limited memory and battery, enabling real-time querying of central health repositories or on-device personal health libraries. This symbiosis has not yet been systematically explored at the systems level, which motivates the architectural synthesis presented here.

3. System Architecture and Infrastructure Trade-offs

Designing a production-grade multi-label retrieval system for wearable biosignals requires navigating a complex landscape of hardware constraints, network availability, privacy regulations, and update cadence. We conceptualize the system as a tiered architecture comprising on-device preprocessing, edge-based hash encoding and caching, and a cloud-hosted primary index. Raw signals are preprocessed locally to extract clean segments and compute instantaneous physiological metrics such as inter-beat intervals and pulse amplitude variability. This preprocessing not only reduces transmission bandwidth but also mitigates privacy risks by removing personally identifiable voice or location data before any signal leaves the device.

The representation learner and hash encoder may be split across tiers depending on the computational budget. A lightweight encoder can reside on the wearable or a companion smartphone, producing compact hash codes that are transmitted to an edge server for initial retrieval against a local cache of frequently accessed health profiles. For exhaustive search across large-scale national biobanks, the cloud tier maintains a distributed index of billions of binary codes, using Hamming distance-based approximate nearest neighbor algorithms that can handle multi-label queries with result diversification. The trade-off between code length and retrieval recall is a central design parameter: shorter codes reduce storage and comparison cost but increase collision probability, potentially merging semantically distinct health states.

Conversely, longer codes improve discriminability but strain the limited memory and battery of wearable devices. System designers must thus calibrate code length not only based on offline benchmark precision-recall curves, but also on field measurements of energy consumption per hash computation and transmission.

Infrastructure choices further intersect with model update strategies. Biosignal distributions drift over time due to seasonal influences, aging populations, and evolving sensor hardware. A static hash encoder will degrade, necessitating frequent retraining. Federated learning offers a privacy-preserving mechanism for aggregating model updates across thousands of devices without centralizing raw biosignals [11]. However, hashing models pose unique challenges for federated optimization: the binary quantization step is non-differentiable, requiring surrogate gradient approximations or alternating optimization schedules. Moreover, ensuring that hash function updates across cohorts do not break backward compatibility of previously stored binary codes in the cloud index demands careful versioning and code translation layers. These are non-trivial systems engineering problems that require co-design between the learning and database communities.

Edge-cloud coordination also needs to respect health data governance frameworks. Many jurisdictions require that personal health data be stored and processed within national borders. A retrieval system may therefore need to deploy region-specific indices while sharing anonymized model updates through differential privacy or secure aggregation. The architecture must be modular enough to allow different parts of the model to be trained on separate data silos: for instance, the statistical-prior module that encodes physiological knowledge could be trained on large public datasets, whereas the fine-tuning of multi-label hash boundaries could occur locally using labeled electronic health records from a single hospital system. Such modularity tests the limits of current interoperability standards, yet it is essential for achieving both high retrieval accuracy and compliance with regulations like the Health Insurance Portability and Accountability Act and the General Data Protection Regulation.

4. Statistical-Prior Guided Representation Learning

The quality of the representation space fundamentally determines the retrieval performance of any downstream hashing mechanism. In wearable biosignal analysis, pure data-driven training often struggles with the long-tailed nature of rare health conditions, susceptibility to confounding artifacts, and poor out-of-distribution generalization. Statistical priors offer a complementary signal that injects domain knowledge directly into the learning process, regularizing the model towards physiologically plausible feature manifolds. Such priors can take multiple forms: explicit generative models of arterial pulse wave propagation that simulate plausible PPG waveform deformations under varying vascular stiffness and heart rate; prior distributions over heart rate variability frequency bands that encode autonomic balance; or noise models grounded in empirical characterization of motion artifacts from accelerometer co-recordings.

A powerful instantiation of this principle involves generative masking architectures that reconstruct randomly corrupted signal segments under the guidance of statistical priors. By training a model to restore missing portions of a biosignal based on learned population-level distributions, the encoder part of the architecture acquires deep structural understanding of pulse morphology, arrhythmic patterns, and their interrelationships. This self-supervised objective can be scaled to millions of unlabeled recordings, producing a foundation model whose intermediate representations serve as input to hash encoders. The resulting representations are robust to sensor dropout and partial occlusion, both common in real-world wearables, because the model has been explicitly trained to infer missing signal fragments from context and prior knowledge.

This approach introduces interesting trade-offs between the strength of the prior and the flexibility left for data-driven discovery. An overly rigid prior, such as a fixed arterial tree model, may discard high-frequency information that carries diagnostic value for conditions like atrial fibrillation with variable ventricular response. Conversely, too weak a prior fails to prevent the model from collapsing onto spurious correlations, such as learning to identify a specific user by noise fingerprint rather than by genuine physiological patterns. The system must therefore incorporate mechanisms to adapt the prior's influence over the course of training, perhaps by annealing the weight of the reconstruction loss or by employing Bayesian frameworks that treat prior strength as a learnable parameter per physiological feature group. The multi-label retrieval downstream task provides a natural validation signal: if retrieval performance on held-out comorbidity categories improves, the prior is beneficial; if it degrades, the prior needs recalibration.

A recent generative masking architecture for PPG signals has demonstrated the viability of this paradigm by integrating statistical-prior informed masking with foundation model pretraining [14]. The model learns to reconstruct photoplethysmography segments through a denoising task that respects known pulse wave characteristics, yielding embeddings that generalize across population subgroups and sensor types. When these embeddings are employed as a substrate for health state retrieval, they significantly improve recall for rare combined conditions compared to representations learned without domain priors. The implications for system design are substantial: pretraining such foundation models requires substantial computational resources, but once trained, they can be distilled into lightweight encoders suitable for on-device hashing, amortizing the initial cost over millions of queries.

5. Deep Hash Encoding for Multi-Label Health State Retrieval

Transforming continuous representations into binary codes for efficient retrieval demands a hashing scheme that preserves the complex similarity structure inherent in multi-label health state assignments. Classical hashing methods rely on pairwise similarity matrices defined by single-label equivalence; extending them to multi-label settings where each instance can be associated with several conditions requires redefining similarity through set overlap measures. Typical choices include the Jaccard index or a weighted sum of partial similarities. Learning hash functions that optimize these multi-label metrics directly is challenging because the Hamming distance between binary codes is discrete and non-differentiable, yet recent advances in continuation methods and asymmetric loss formulation have made such optimization possible.

A particularly effective strategy, rooted in self-supervised semantic excavation and margin-scalable constraints, constructs hash codes by alternating between a feature extraction backbone and a set of auxiliary tasks that excavate latent semantic structures without exhaustive pair labels [15]. The approach leverages the inherent asymmetry between the high-dimensional continuous representation space and the low-dimensional binary space by introducing differentiable surrogates that relax the binary constraints during training while enforcing strict quantization during inference. The margin-adaptive component dynamically adjusts the desired distance between codes of dissimilar health states based on the degree of label overlap, ensuring that partially overlapping comorbidity profiles are mapped to codes with intermediate Hamming distances rather than being forced into a binary similar-or-dissimilar decision. This nuanced encoding is crucial for clinical retrieval, where a query from a patient with diabetes, hypertension, and mild sleep apnea should retrieve cases that share a substantial subset of these conditions, even if the overlap is not perfect.

The application of such hashing techniques to biosignal retrieval surfaces several system-level considerations. The first is code length selection. In multimedia retrieval, code lengths of 128 or 256 bits are common, but biosignal data may require longer codes to disambiguate the finer granularity of health states. Pilot experiments on photoplethysmography datasets indicate that 512-bit codes offer a reasonable balance between retrieval accuracy and memory footprint for millions of records. However, the energy cost of computing a 512-bit code on a microcontroller via a compact convolutional network must be quantified and compared against the energy cost of transmitting the raw representation. In scenarios where edge inference is performed on a smartphone, longer codes are acceptable; for always-on, battery-constrained wristbands, code length must shrink, possibly relying on hierarchical hashing where a short code pre-filters candidates and a longer code is computed on demand for the top matches.

A second consideration is the integration of temporal dynamics. Biosignals are inherently sequential, and a single recording may contain evolving health states. A hash encoder that processes a fixed-length window will produce a code that mixes information from multiple phases, potentially obscuring transient events such as paroxysmal arrhythmias. System designers may therefore adopt multi-scale hashing, where multiple codes are generated from overlapping or segmented windows, and retrieval is performed with a temporal aggregation strategy that weights window-level matches according to their temporal proximity or clinical relevance. This adds a layer of complexity to the index structure but substantially improves retrieval of episodic conditions.

The interaction between statistical-prior guided representation and hashing also deserves scrutiny. If the representation space is highly structured due to strong priors, the hash function may converge more quickly and achieve better retrieval performance with fewer labeled examples. Conversely, the discrete nature of hashing can mask subtle improvements in representation quality, as small embedding changes may not flip hash bits. This calls for an evaluation protocol that measures retrieval performance not only by Hamming ranking metrics but also by label set overlap metrics at various operating points, together with calibration of retrieval confidence.

6. Robustness, Fairness, and Energy Sustainability

Deploying multi-label retrieval in real-world healthcare settings demands far more than high average precision. Robustness to naturally occurring distribution shifts, fairness across demographic groups, and energy sustainability form a triad of non-functional requirements that must be embedded into the system architecture from the outset. Biosignal morphology varies significantly across skin tones, age groups, body mass indices, and the presence of peripheral vascular diseases. A retrieval system trained predominantly on data from lighter-skinned individuals in resting conditions may produce degraded representations for darker-skinned users or during physical activity, leading to differential retrieval performance that can perpetuate health disparities. Such biases are not hypothetical: pulse oximetry, which shares the same underlying photoplethysmography principle, has well-documented racial biases in oxygen saturation estimation, a fact that underscores the need for deliberate mitigation strategies in any downstream AI system.

Addressing fairness in the multi-label retrieval setting requires a combination of dataset curation, algorithmic debiasing, and evaluation methodology. During pretraining, data acquisition protocols must ensure balanced representation across skin phototypes, sensor types, and activity states. Statistical priors can serve as a double-edged sword: if the prior distributions are estimated from biased population samples, they will reinforce demographic skews. Consequently, prior construction should incorporate stratified population models, and the generative masking process should be audited to ensure that reconstruction quality is uniform across groups. At the hash encoding stage, fairness constraints can be integrated into the loss function to penalize hash code distributions that exhibit significant demographic predictability, effectively learning codes that are invariant to protected attributes while preserving health state semantics. At the infrastructure level, continuous monitoring of retrieval quality stratified by demographic and device metadata must be embedded into the system's telemetry, with automated triggers for model recalibration when disparities exceed predefined thresholds.

Energy sustainability is an equally pressing concern. The cumulative carbon footprint of training large foundation models has attracted scrutiny, yet the inference-time energy cost of retrieval is arguably more critical for wearables, as it directly impacts device battery life and user adherence. The trade-off between model complexity and power consumption can be navigated through a cascade of optimization techniques: knowledge distillation from a large teacher model to a compact student encoder; quantization-aware training that reduces numerical precision; and neural architecture search tailored to the specific hardware accelerator of the wearable system-on-chip. Deep hashing contributes positively to energy efficiency because Hamming distance computation is extremely cheap compared to floating-point inner products, and the compressed code reduces memory access energy. Nevertheless, the total energy must be holistically minimized, accounting for signal preprocessing, hash computation, and wireless transmission of the query code. System-level profiling on representative hardware platforms is indispensable to validate that the theoretical efficiency gains translate into practical battery savings.

7. Deployment, Governance, and Policy Implications

Bringing a multi-label health state retrieval system into clinical workflows requires navigating an intricate regulatory landscape. In the United States, the Food and Drug Administration classifies software that analyzes physiological signals for diagnostic purposes as a medical device, subjecting it to premarket review and postmarket surveillance. A retrieval system that presents similar cases and aggregated health states to a clinician without autonomously rendering a diagnostic decision may be positioned as clinical decision support software, potentially qualifying for less burdensome regulatory pathways if it meets non-device criteria. However, the distinction becomes blurred if the retrieval output is integrated into automated triage or alerting systems. The system's designers must work closely with regulatory experts to craft an intended use statement that matches the level of autonomy to the appropriate risk classification.

Transparency and explainability are pivotal governance requirements. When a health state is inferred from the labels of retrieved neighbors, the clinician must be able to inspect which specific cases contributed to that inference and why they were considered similar. This demands an explainable retrieval interface that visualizes the matches in terms of both raw signal segments and hash code proximity, while also providing an uncertainty measure that conveys the density of the retrieved neighborhood. The hash codes themselves can be structured in a hierarchical fashion, where coarser bits correspond to broad physiological categories and finer bits capture subtle distinctions, thus offering a human-interpretable navigation of the manifold even in encoded form.

Data governance extends beyond privacy to encompass data sovereignty, especially when models are deployed across national borders. The architecture described in Section 3 enables a federated index where each jurisdiction hosts its own hash database, and queries are routed accordingly. Such a design must comply with cross-border data transfer restrictions, which may require that only anonymized hash codes, not raw signals, cross jurisdictional boundaries. The representation learning pipeline must therefore be evaluated for the risk of re-identification from hash codes, particularly given advances in inversion attacks that can reconstruct inputs from intermediate representations. Differential privacy guarantees can be applied during hash computation, though they introduce a delicate trade-off between privacy budget and retrieval precision. These are not merely technical choices but policy decisions that require multi-stakeholder deliberation involving clinicians, patients, ethicists, and legal experts.

Looking forward, the emergence of universal health data interoperability standards, such as HL7 Fast Healthcare Interoperability Resources profiles for wearables, creates an opportunity to embed multi-label retrieval as a microservice within broader health information exchanges. Standardization of hash code formats, retrieval APIs, and audit logs would allow different vendors and health systems to share indices while maintaining data locality and security. Such an ecosystem could amplify the benefits of large-scale retrieval: a rare disease patient in a small clinic could be matched against a global index of anonymized cases, dramatically accelerating diagnostic odyssey. The governance framework would need to evolve accordingly, perhaps through international consortia that define acceptable use policies and liability distribution among index contributors, model maintainers, and deploying institutions. These institutional innovations are as critical as the technical ones to realizing the vision of equitable, efficient health state retrieval.

8. Conclusion

Multi-label health state retrieval from wearable biosignals sits at the intersection of representation learning, information retrieval, and health informatics, and its full realization demands a systems perspective that interlocks algorithmic innovation with infrastructure design, fairness engineering, and policy alignment. This paper has articulated how coupling statistical-prior guided representation learning with deep hash encoding can yield a retrieval pipeline that is both semantically rich and computationally efficient, while surfacing the structural trade-offs that define its practical viability. Statistical priors, whether encoded through generative masking or physiological models, imbue representations with robustness and generalization that extend to rare and comorbid conditions. Deep hashing, particularly methods employing self-supervised semantic excavation and margin-adaptive constraints, compresses these representations into binary codes that enable sublinear retrieval without sacrificing the nuanced multi-label similarity structure essential for clinical utility.

We have underscored that no component of this pipeline can be optimized in isolation. The choice of prior strength influences the representation manifold, which in turn affects the optimal code length and the convergence of federated hash learning. The hardware platform dictates feasible model complexity, which constrains the fidelity of learned priors. The regulatory framework shapes what kind of decision support is permissible, which must be reflected in the retrieval interface and output format. Cross-cutting concerns of fairness and energy sustainability must be monitored continuously and fed back into the design loop, making the system a living socio-technical entity rather than a static product.

Future work should pursue several directions. From an engineering standpoint, end-to-end joint optimization of the representation, hashing, and indexing layers through hardware-aware neural architecture search could yield Pareto-optimal configurations for different deployment scenarios. From a clinical perspective, prospective studies that measure the impact of multi-label retrieval on diagnostic accuracy, clinical workflows, and patient outcomes are essential to move beyond benchmark metrics. From a governance angle, the development of audit protocols and certification standards for retrieval-based decision support systems will be crucial to earning public trust. The convergence of these efforts promises a new generation of wearable analytics that do not merely count steps or detect singular events, but holistically retrieve and interpret the multiplicity of human health states, responsibly and at scale.

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. https://doi.org/10.1038/s41591-018-0268-3

2. Liang, Y., Chen, Z., Liu, G., & Elgendi, M. (2018). A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China. Scientific Data, 5, 180020. https://doi.org/10.1038/sdata.2018.20

3. Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., ... & Ribeiro, A. L. P. (2020). Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications, 11, 1760. https://doi.org/10.1038/s41467-020-15432-4

4. 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. 5607–5618). PMLR.

5. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2023). TimesNet: Temporal 2D-variation modeling for general time series analysis. In International Conference on Learning Representations.

6. Sarkar, P., & Etemad, A. (2020). Self-supervised ECG representation learning for emotion recognition. In 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 1234–1238). IEEE.

7. Zhu, H., Long, M., Wang, J., & Cao, Y. (2016). Deep hashing network for efficient similarity retrieval. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 2415–2421).

8. Liu, W., Wang, J., Ji, R., Jiang, Y. G., & Chang, S. F. (2012). Supervised hashing with kernels. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2074–2081). IEEE.

9. Chen, Y., Lai, Z., Ding, Y., Lin, K., & Wong, W. K. (2020). Deep supervised hashing with anchor graph for large-scale image retrieval. IEEE Transactions on Image Processing, 29, 179–194. https://doi.org/10.1109/TIP.2019.2927632

10. Rajkomar, A., Hardt, M., Howell, M. D., Corrado, G., & Chin, M. H. (2018). Ensuring fairness in machine learning to advance health equity. Annals of Internal Medicine, 169(12), 866–872. https://doi.org/10.7326/M18-1990

11. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. npj Digital Medicine, 3, 119. https://doi.org/10.1038/s41746-020-00323-1

12. Wan, J., Al-awlaqi, M., Li, M., O'Grady, M., Gu, X., Wang, J., & Cao, N. (2018). Wearable IoT enabled real-time health monitoring system. IEEE Access, 6, 4248–4258. https://doi.org/10.1109/ACCESS.2017.2788297

13. He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. Nature Medicine, 25(1), 30–36. https://doi.org/10.1038/s41591-018-0307-0

14. Guo, Z., Chen, T., Jiao, Y., Pan, Y., Hu, X., & Ferrario, M. (2026). SIGMA-PPG: Statistical-prior Informed Generative Masking Architecture for PPG Foundation Model. arXiv preprint arXiv:2601.21031.

15. Yu, Z., Wu, S., Dou, Z., & Bakker, E. M. (2022). Deep hashing with self-supervised asymmetric semantic excavation and margin-scalable constraint. Neurocomputing, 483, 87-104.

16. Guan, Y., & Plötz, T. (2017). Ensembles of deep LSTM learners for activity recognition using wearables. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1(2), 28. https://doi.org/10.1145/3090076

17. Han, S., Mao, H., & Dally, W. J. (2016). Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding. In International Conference on Learning Representations.

18. 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. https://doi.org/10.1109/MSP.2020.2975749

19. Schmidt, P., Reiss, A., Dürichen, R., & Laerhoven, K. V. (2018). Introducing WeSAD, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM International Conference on Multimodal Interaction (pp. 400–408). ACM.

20. Tison, G. H., Sanchez, J. M., Ballinger, B., Singh, A., Olgin, J. E., Pletcher, M. J., ... & Marcus, G. M. (2019). Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiology, 4(6), 610–617. https://doi.org/10.1001/jamacardio.2019.1329

21. Mao, Y., You, C., Zhang, J., Huang, K., & Letaief, K. B. (2017). A survey on mobile edge computing: The communication perspective. IEEE Communications Surveys & Tutorials, 19(3), 1622–1657. https://doi.org/10.1109/COMST.2017.2702372

22. Liang, Z., Chapa-Martell, M., & Cakmak, A. (2021). An FHIR-based framework for integrating wearable sensor data into electronic health records. JMIR Medical Informatics, 9(1), e23399. https://doi.org/10.2196/23399

23. Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E., & Valley, T. S. (2020). Racial bias in pulse oximetry measurement. New England Journal of Medicine, 383, 2477–2478. https://doi.org/10.1056/NEJMc2029240

24. Yue, Y., Khanal, A., Lyu, T., Weissman, S., & Liang, C. (2025, May). EHR Phenotyping Methods for Measuring Treatment Adherence Among People Living With HIV in All of Us: Towards Disparities and Inequalities in HIV Care Continuum. In AMIA Annual Symposium Proceedings (Vol. 2024, p. 1294).