Foundation Model–Driven Multimodal Health Retrieval: Integrating PPG Physiological Signals and Medical Imaging via Scalable Deep Hashing

Aarav Chopra
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
aarav.chopra395@uab.edu

Sunil J. Gandhi
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
sunil.gandhi197@binghamton.edu

Yuetang Pan
School of Information Technology, University of Cincinnati, Cincinnati, OH, USA.
yuetang.work@uc.edu

Abstract

The rapid proliferation of wearable sensors and medical imaging has created unprecedented volumes of heterogeneous health data, motivating the development of intelligent retrieval systems that can exploit both continuous physiological signals and visual diagnostic evidence. This paper presents a system-level investigation of foundation model–driven multimodal health retrieval that integrates photoplethysmogram (PPG) signals and medical imaging through scalable deep hashing. We argue that the confluence of self-supervised foundation models pre-trained on diverse physiological and imaging corpora, coupled with advanced asymmetric deep hashing techniques, offers a transformative pathway toward cross-modal semantic search in clinical environments. Departing from task-specific model training, the proposed architecture leverages modality-specific foundation models to produce unimodal embeddings that are then aligned in a shared latent space and encoded into compact binary hash codes via margin-scalable, self-supervised hashing mechanisms. The paper examines critical system dimensions including modular design, infrastructure requirements, computational sustainability, robustness against distributional shifts, fairness across demographic groups, and governance frameworks that reconcile performance with patient privacy and regulatory compliance. By synthesizing insights from medical AI, hashing theory, and socio-technical systems research, we delineate the structural trade-offs inherent in building a production-grade multimodal retrieval system for health analytics. The discussion further elaborates on deployment strategies that span edge–cloud continuua, the necessity of continuous model monitoring, and the policy implications of embedding such retrieval engines within clinical decision support pipelines. The paper advances a forward-looking perspective on how foundation model–driven hashing can unlock new classes of health applications while remaining attentive to the ethical and operational complexities of large-scale healthcare AI.

Keywords

foundation model, multimodal health retrieval, photoplethysmography, medical imaging, deep hashing, self-supervised learning, scalable systems, fairness, governance.

1. Introduction

Contemporary healthcare ecosystems generate a vast and heterogeneous mixture of data modalities, encompassing high-resolution medical images such as chest radiographs, computed tomography scans, and magnetic resonance imaging, alongside continuous physiological waveforms captured by wearable devices, among which photoplethysmogram (PPG) signals have become especially pervasive. These two information streams capture complementary aspects of patient physiology: imaging provides spatially localized structural and pathological evidence, while PPG time series reflect systemic cardiovascular dynamics, autonomic regulation, and hemodynamic fluctuations. The ability to retrieve clinically relevant information across these modalities, for instance, to find medical imaging studies that share physiological signatures similar to a given PPG recording or vice versa, promises to enrich diagnostic reasoning, support differential diagnosis, and accelerate retrospective case review. Realizing such cross-modal retrieval at scale, however, confronts fundamental challenges of semantic alignment, representation learning, storage efficiency, and query speed, which demand a system architecture that is simultaneously expressive, compact, and robust. Recent advances in foundation models and deep hashing have independently transformed representation learning in health data analysis, yet their systematic integration for multimodal retrieval remains underexplored from a systems perspective. This paper examines the vision of a foundation model–driven multimodal health retrieval platform that integrates PPG physiological signals and medical imaging via scalable deep hashing, focusing on architectural design, infrastructure trade-offs, and the broader governance dimensions that accompany deployment in real-world clinical environments.

The momentum behind foundation models in medicine stems from their capacity to learn transferable representations from large-scale, unlabeled corpora, thereby reducing the dependency on task-specific annotated datasets and improving generalization across downstream settings [1], [2]. Meanwhile, deep hashing has evolved into a cornerstone technique for approximate nearest neighbor search in high-dimensional multimedia databases, enabling sub-linear retrieval times and drastic memory reductions through binary code representations [3]. The marriage of these two paradigms is particularly compelling for multimodal health retrieval, where the intrinsic complexity of each modality demands powerful feature extractors, and the sheer volume of data produced by modern health systems necessitates retrieval efficiency that goes beyond brute-force comparison. The design of such a system must account for the asymmetries between modalities—PPG signals are temporally dense, low-dimensional, and prone to motion artifacts and inter-subject variability, whereas medical images are highly structured, high-dimensional, and exhibit domain-specific anatomical priors—while preserving semantic cross-modal associations that are clinically actionable. The central thesis of this work is that a modular architecture, built from pre-trained modality-specific foundation models and connected by a shared hashing layer optimized through self-supervised asymmetric constraints, can deliver retrieval performance that balances accuracy, scalability, and interpretability, provided that system-level concerns such as data governance, model fairness, and deployment sustainability are treated as first-class design requirements.

2. Background and Related Work

The analysis of PPG signals has matured from simple peak detection for heart rate estimation to deep learning–based extraction of a rich set of cardiovascular and respiratory parameters. Foundational studies established the physiological underpinnings of the PPG waveform and its susceptibility to peripheral perfusion changes, sensor placement, and ambient conditions [4]. Beyond basic vital sign estimation, modern neural architectures have demonstrated the capacity to infer blood pressure, arterial stiffness, and even early signs of cardiac arrhythmia from raw PPG traces, prompting interest in the use of these signals as a low-cost, continuous monitoring modality in both clinical and consumer settings. In parallel, medical imaging retrieval has benefited from decades of research in content-based image retrieval, which has recently been reinvigorated by deep convolutional networks and vision transformers that encode rich semantic features capable of capturing subtle pathological patterns. The emergence of cross-modal hashing methods designed to map heterogeneous data types into a common Hamming space has enabled large-scale similarity search across modalities such as images and text captions [5], yet the translation of these techniques to physiological signals combined with imaging remains rare due to the marked differences in data characteristics and the scarcity of paired multimodal health datasets.

Self-supervised learning has become a pivotal enabler for health representation learning, reducing reliance on manual annotation by exploiting inherent data structure. In medical imaging, contrastive frameworks that maximize agreement between differently augmented views of the same image have yielded representations that rival or surpass those from supervised pre-training on extensive labeled datasets [6]. For physiological time series, the CLOCS framework demonstrated that contrastive learning over cardiac signals recorded across varied spatial contexts and patients can produce versatile representations transferable to diverse downstream tasks, including PPG-based arrhythmia detection and sleep staging [7]. The proliferation of these self-supervised models has paved the way for the concept of medical foundation models—large, general-purpose models that can be fine-tuned for a spectrum of tasks without extensive retraining—and sparked a series of works examining their applicability across radiology, pathology, and wearable sensor analytics [2], [8]. Despite these advances, the majority of foundation model research in health has remained within a single modality, leaving the challenge of building cross-modal embedding spaces that coherently represent physiological signals and images largely unresolved.

Deep hashing has been extensively studied in the retrieval community, with supervised, semi-supervised, and unsupervised variants developed to generate binary codes of controllable length while preserving semantic neighborhood structures. Recent work has introduced self-supervised asymmetric semantic excavation and margin-scalable constraint mechanisms that jointly optimize the hash code generation and the semantic similarity ranking without requiring exhaustive pairwise labeling, achieving state-of-the-art retrieval precision on multimedia benchmarks [9]. These techniques are attractive for health applications because the pairwise supervision costs across modalities would be prohibitive; instead, leveraging asymmetric learning objectives that exploit the natural coherence between a PPG segment and a corresponding diagnostic image within the same patient encounter can drive the hashing network toward clinically meaningful codebooks. Furthermore, medical vision-language models, such as those trained on paired radiology images and reports using contrastive objectives, have illustrated that semantically rich cross-modal alignments can be learned even when the modalities differ radically in structure, providing a conceptual benchmark for the PPG–imaging alignment problem [10]. Collectively, these lines of work establish the

technical foundations upon which a multimodal health retrieval system can be constructed, while also highlighting the gaps that a dedicated system-level design must fill.

3. System Architecture and Design Philosophy

The proposed system adopts a modular, three-stage architecture that decouples modality-specific feature extraction, cross-modal alignment, and hash code generation. In the first stage, separate foundation models pre-trained on large-scale PPG and medical imaging corpora serve as frozen or partially fine-tuned encoders, transforming each raw signal and image into a dense embedding vector. For PPG, a temporal transformer-based architecture pre-trained on multi-source wearable datasets using masked reconstruction and contrastive objectives can capture both short-term waveform morphology and long-term heart rate variability patterns, while for imaging, a vision transformer pre-trained on millions of radiographs using self-supervised techniques analogous to those described in large-scale medical image analysis studies provides a strong visual backbone [6], [11]. The second stage projects these unimodal embeddings into a shared latent space through modality-specific projection heads and aligns the representations using a combination of contrastive loss and cycle-consistency constraints derived from patient-level metadata that naturally link PPG recordings to imaging studies obtained during the same clinical episode. The alignment module is trained to ensure that semantically related PPG segments and images are mapped to nearby regions while preserving intra-modal discriminability, a balance that is critical for downstream retrieval relevance.

The third stage introduces a hash layer that compresses the shared representations into compact binary codes of configurable length, typically ranging from 64 to 256 bits, through a combination of quantization-aware training and similarity-preserving optimization. The design draws inspiration from deep hashing frameworks that employ self-supervised asymmetric constraints, where the network learns to predict the similarity ranking between a query code and a database of binary codes without requiring exhaustive pairwise supervision [9]. By adopting a margin-scalable objective, the system can tune the trade-off between recall and precision according to clinical deployment requirements: a narrower margin yields higher precision for diagnostic filtering tasks, while a wider margin improves recall for case retrieval during research or educational review. The entire architecture is trained end-to-end after the initial foundation models have been pre-trained, using a curated multi-institutional dataset of paired PPG and imaging records, with careful attention to data partitioning strategies that prevent patient-level leakage across training, validation, and testing splits. This modular design confers substantial engineering benefits: the foundation models can be updated independently as new pre-training corpora become available, the alignment module can be refined with additional paired data without altering the hash code infrastructure, and the hashing layer can be replaced or scaled to different bit lengths to meet evolving latency and storage constraints, supporting a system life cycle that accommodates continuous improvement.

From a governance standpoint, the architecture must incorporate auditing checkpoints at each stage. The foundation models, which may have been trained on data from diverse demographic and geographic populations, require bias audits to detect differential performance across subgroups before they are integrated into the alignment pipeline [13]. The alignment module, in turn, must be evaluated for cross-modal fairness, ensuring that retrieval quality does not degrade systematically for patients with certain PPG signal characteristics—such as darker skin tones known to affect optical sensor accuracy—or for specific imaging

modalities. The hash function adds an additional layer of opacity, as binary codes are inherently less interpretable than continuous embeddings, necessitating post-hoc explanation techniques that can project influential input regions onto the original signal or image to support clinician trust. These concerns must be embedded within the development process rather than treated as afterthoughts, echoing the principles laid out in recent analyses of foundation model risks and the importance of transparent documentation practices [12]. Furthermore, the system must respect data minimization standards, processing only the necessary signals and images for the intended retrieval task and supporting mechanisms for patient consent withdrawal that propagate through the hashing index.

4. Foundation Model Adaptation for PPG and Imaging Modalities

The effectiveness of the retrieval system hinges on the quality and breadth of the foundation models that serve as upstream encoders. For PPG, the pre-training strategy must contend with the high inter-subject variability, sensor noise, and limited availability of labeled clinical endpoints. A self-supervised framework that combines masked signal modeling, where random segments of the waveform are masked and predicted from the surrounding context, with contrastive objectives that treat recordings from the same individual under similar conditions as positive pairs, has shown promise in learning representations that generalize to tasks such as blood pressure estimation and atrial fibrillation detection [16]. The statistical-prior informed generative masking architecture proposed for PPG foundation models introduces a principled approach to incorporate domain knowledge about PPG signal periodicity and pulse morphology into the pre-training pipeline, enhancing representation robustness even when fine-tuning data is scarce [17]. Such a foundation model can be trained on aggregated datasets spanning diverse wearable device models, sampling rates, and clinical populations, ensuring that the resulting embeddings capture a broad spectrum of cardiovascular physiology. The computational cost of pre-training these models, however, raises sustainability concerns that require careful accounting; training a PPG foundation model on hundreds of thousands of hours of recordings can incur substantial energy consumption and carbon emissions, which must be weighed against the downstream clinical utility and considered when deciding between building new models from scratch versus adapting existing open-source pre-trained weights [20].

Medical imaging foundation models have been trained using massive collections of radiographs, computed tomography volumes, and magnetic resonance imaging studies, often leveraging variations of DINO, MAE, or CLIP-style objectives to capture both local and global visual features. The use of vision transformers with large patch sizes and global self-attention has proven particularly effective for chest radiography, where diffuse disease patterns demand holistic image understanding. When integrating an imaging foundation model into the multimodal retrieval system, a critical design choice concerns the layer from which embeddings are extracted—whether from the final classification token, intermediate feature maps, or a concatenation of multi-scale representations—and the degree of fine-tuning allowed. Freezing the backbone encoder preserves generalization and reduces the risk of catastrophic forgetting, but may limit the model's ability to adapt to the specific visual characteristics of the retrieval task, such as the need to emphasize subtle interstitial patterns relevant to matching with PPG-derived respiratory indices. Partial fine-tuning of the later transformer layers, combined with a low-rank adaptation strategy, can offer a pragmatic compromise that balances adaptation and stability, and has been shown to maintain robustness in medical imaging applications under distribution shift [6]. The allied concern of

interpretability becomes salient here: when a retrieved image contributes to a clinical decision, the system must be capable of generating saliency maps that highlight the anatomical regions most responsible for the match, using methods such as SHAP or integrated gradients, thereby enabling clinicians to verify that the retrieval is based on medically plausible features rather than spurious correlations [18].

5. Scalable Deep Hashing for Cross-Modal Retrieval

The hash code generation module sits at the core of the system's scalability, transforming high-dimensional cross-modal embeddings into compact binary strings that can be compared using Hamming distance in constant time and stored with minimal memory overhead. The choice of code length represents a pivotal architectural trade-off: shorter codes reduce storage and accelerate retrieval but compress the embedding space more aggressively, potentially collapsing semantically distinct clinical patterns into identical hash buckets, while longer codes preserve finer-grained similarity at the cost of larger index sizes and higher query latency. The margin-scalable constraint mechanism allows the system to push the decision boundary governing hash code similarity to adapt to the precision–recall demands of specific clinical use cases. For applications such as retrieving historical cases with a known PPG phenotype to inform diagnosis, a recall-oriented setting with a relaxed margin may be appropriate; for screening tasks where false positives incur substantial downstream cost, a high-precision regime with tightened margin is preferred. This tunability is achieved without retraining the entire network, simply by adjusting the margin hyperparameter at inference time, a property that significantly enhances the operational flexibility of the system.

Training the hashing module with self-supervised asymmetric objectives circumvents the need for exhaustive pairwise labels across imaging and PPG modalities. The asymmetric formulation uses the high-dimensional continuous embeddings of one modality as a supervisory signal for the binary codes of the other, exploiting the fact that images and PPG signals acquired from the same patient encounter share underlying physiological states. Through a carefully designed sampling strategy that constructs positive pairs from temporally aligned patient data and negative pairs through in-batch random sampling and hard negative mining, the hash function learns to position codes such that semantically consistent cross-modal pairs yield low Hamming distances. The resulting binary codes can be indexed using multi-index hashing or quantization-based search structures that support sub-linear time queries even for databases containing millions of records, enabling the system to scale horizontally by distributing shards of the hash tables across a cluster of servers. To ensure that retrieval remains reliable as the underlying data distribution evolves—such as when new sensor models or imaging protocols are introduced—the system must incorporate periodic re-indexing and incremental learning cycles that update the hash function without requiring complete model retraining, a process that can be facilitated by federated averaging techniques that aggregate model updates from multiple clinical sites without sharing raw patient data [19].

6. Infrastructure, Deployment, and Sustainability

Deploying a foundation model–driven multimodal health retrieval system in a real-world clinical environment demands careful consideration of the infrastructure topology that connects edge devices, institutional servers, and cloud platforms. Wearable PPG sensors typically operate on resource-constrained embedded systems, where transmitting raw signals to a central server for feature extraction may violate latency and privacy requirements. A tiered architecture that performs lightweight on-device pre-processing, such as signal quality assessment and basic feature extraction using a compressed student model distilled from the

large foundation encoder, can reduce upstream bandwidth consumption and enable local caching of frequently retrieved cases. The computationally intensive foundation model inference and hashing operations can then be executed on hospital-edge servers equipped with GPU accelerators, while the global index and model registry reside in a secure cloud environment that orchestrates federated fine-tuning rounds and maintains the master hash tables. This hybrid edge–cloud design balances the benefits of low-latency inference at the point of care with the economies of scale afforded by centralized model management, though it introduces network reliability concerns that must be mitigated through offline fallback modes and asynchronous update protocols.

Sustainability considerations have become increasingly prominent in the design of large-scale AI systems for health. The carbon footprint associated with pre-training PPG and imaging foundation models can be substantial, requiring transparent reporting of energy consumption and a commitment to using carbon-efficient data centers and renewable energy sources where possible. Model compression techniques, such as quantization, pruning, and knowledge distillation, can shrink the inference-time footprint of the deployed encoders, while the hashing module inherently reduces the storage energy costs associated with nearest neighbor search by several orders of magnitude compared to dense embedding retrieval. A lifecycle assessment of the system should weigh the one-time training carbon expenditure against the ongoing operational savings and the potential clinical value generated, a calculus that remains underdeveloped in the medical AI literature and calls for standardized sustainability benchmarks analogous to those used in general machine learning [20]. Institutional procurement policies for AI infrastructure may increasingly factor in these environmental metrics, creating a feedback loop that incentivizes the development of energy-proportional and computationally efficient model architectures.

Data governance infrastructure must be tightly integrated with the retrieval pipeline to satisfy the stringent requirements of health data protection regulations such as HIPAA and GDPR. The system should implement attribute-based access control, ensuring that only authorized personnel can initiate queries and that the scope of retrievable records respects patient consent directives and institutional data-sharing agreements. An audit trail logging every retrieval request, together with the hash codes and the returned case identifiers, provides a mechanism for post-hoc accountability and supports forensic analysis in the event of a privacy breach. Furthermore, the capability to unlearn a patient's data—that is, to remove their influence from the hash index and foundation model parameters without requiring full retraining—is an emerging requirement that intersects with both machine unlearning research and the broader policy landscape around the right to be forgotten. While exact unlearning in deep hashing systems remains an open research problem, approximate strategies that regenerate hash tables after targeted data deletion and fine-tune the model on a cleansed corpus offer a pragmatic interim solution.

7. Robustness, Fairness, and Policy Implications

The clinical deployment of a multimodal health retrieval system must contend with a spectrum of robustness challenges. PPG signals are susceptible to motion artifacts, sensor displacement, and ambient light interference, all of which can degrade the quality of the extracted features and propagate errors into the hash code generation. The foundation model's pre-training on diverse, noisy data provides a degree of inherent robustness, but the system should additionally incorporate a signal quality index that gates whether a query is submitted to the retrieval engine, preventing low-confidence matches from reaching the clinician.

Adversarial robustness is another concern: carefully crafted perturbations imperceptible to the human eye could potentially cause the hash function to produce incorrect codes, leading to retrieval of irrelevant or deliberately misleading medical images. While adversarial attacks on multimodal health retrieval have not been extensively demonstrated in practice, threat modeling and stress testing using projected gradient descent attacks on the input space should be part of the system validation protocol, and certified robustness techniques could be explored to provide formal guarantees for safety-critical applications.

Fairness across demographic groups remains one of the most pressing ethical challenges in health AI, and the multimodal retrieval system introduces unique fairness dimensions. The optical properties of PPG sensors cause variations in signal quality as a function of skin pigmentation, which can lead to systematically different embedding distributions for patients with darker skin tones, potentially biasing retrieval results toward cases from lighter-skinned populations if not explicitly mitigated [13]. Similarly, medical imaging datasets have well-documented imbalances in disease prevalence and image quality across racial, ethnic, and socioeconomic groups, which the imaging foundation model may inadvertently encode. Addressing these disparities requires a multi-pronged strategy: pre-training datasets must be curated to maximize demographic representation, bias audits must be performed on the aligned embedding space and the hash code distribution using fairness metrics such as equalized odds and demographic parity, and post-processing calibration techniques that adjust the Hamming distance threshold on a per-group basis can be applied to equalize retrieval quality. Transparency reports that disclose the demographic composition of the training data and the measured fairness metrics for each release of the system would align with emerging best practices in responsible AI deployment and support meaningful external oversight.

The policy landscape surrounding multimodal health retrieval systems is still taking shape. In the United States, the Food and Drug Administration is advancing regulatory frameworks for software as a medical device that incorporate continuous learning algorithms, raising questions about how frequently a retrieval system can be updated before it requires re-certification. In the European Union, the AI Act classifies many healthcare AI applications as high-risk, imposing requirements for human oversight, data quality, and technical documentation that would directly apply to a production retrieval system. Beyond regulatory compliance, the clinical adoption of such a system depends on building trust with healthcare professionals through rigorous prospective validation studies that demonstrate not only retrieval accuracy but also improvements in clinical outcomes, reductions in diagnostic turnaround time, and cost-effectiveness. Incorporating clinician feedback loops into the design, where retrieved cases can be rated for relevance and these ratings used to refine the hashing model through active learning cycles, can both improve system performance and foster a sense of shared ownership that is essential for sustainable integration into clinical workflows. The policy discourse must also grapple with the tension between the benefits of open-source model sharing—which accelerates innovation and reproducibility—and the risks of bad actors exploiting publicly available foundation models to reverse-engineer sensitive health signatures, suggesting that controlled-access model repositories and usage agreements may become normative in this domain.

8. Conclusion

This paper has articulated a system-level vision for foundation model–driven multimodal health retrieval that unifies PPG physiological signals and medical imaging through scalable deep hashing, framing the technical, infrastructural, and governance dimensions as an

integrated design challenge. By coupling modality-specific foundation models with a cross-modal alignment module and a self-supervised asymmetric hashing layer, the proposed architecture offers a path toward semantically rich, computationally efficient retrieval capable of operating across the heterogeneous data modalities that define modern healthcare. The modular design philosophy supports incremental model updates, configurable retrieval precision–recall trade-offs, and compatibility with tiered edge–cloud deployment models, while the explicit incorporation of bias monitoring, interpretability, and data governance mechanisms reflects a socio-technical orientation that treats ethical robustness as a system property rather than an external constraint. Critical forward-looking challenges remain, including the need for large-scale, ethically curated multimodal health datasets, the development of standardized benchmarks for cross-modal health retrieval, and the advancement of machine unlearning techniques suited to hashing-based indexing structures. As wearable sensing becomes ubiquitous and medical imaging archives continue to expand, the ability to navigate the latent connections across these data streams will become an increasingly valuable clinical capability, and the architectural principles outlined here provide a conceptual foundation upon which such capabilities can be responsibly realized.

References

1. Huang, S. C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. npj Digital Medicine, 3(1), 136.

2. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259–265.

3. Wang, J., Zhang, T., Song, J., Sebe, N., & Shen, H. T. (2018). A survey on learning to hash. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4), 769–790.

4. Elgendi, M. (2012). On the analysis of fingertip photoplethysmogram signals. Current Cardiology Reviews, 8(1), 14–25.

5. Jiang, Q. Y., & Li, W. J. (2017). Deep cross-modal hashing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3232–3240).

6. Azizi, S., Mustafa, B., Ryan, F., Hénaff, O. J., Buck, J. R., & others. (2021). Big self-supervised models advance medical image classification. Nature Biomedical Engineering, 5, 512–523.

7. Kiyasseh, D., Zhu, T., & Clifton, D. A. (2021). CLOCS: Contrastive learning of cardiac signals across space, time, and patients. In International Conference on Machine Learning (pp. 5606–5615). PMLR.

8. Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A., ... & Topol, E. (2021). Deep learning-enabled medical computer vision. npj Digital Medicine, 4(1), 5.

9. 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.

10. Zhang, Y., Jiang, H., Miura, Y., Manning, C. D., & Langlotz, C. P. (2022). Contrastive learning of medical visual representations from paired images and text. In Machine Learning for Healthcare Conference (pp. 2–25). PMLR.

11. Liu, H., Wang, R., Shan, S., & Chen, X. (2016). Deep supervised hashing for fast image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2064–2072).

12. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

13. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

14. Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., ... & Braren, R. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. Nature Machine Intelligence, 3(6), 473–484.

15. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38.

16. Slapničar, G., Mlakar, N., & Luštrek, M. (2019). Blood pressure estimation from photoplethysmogram using a spectro-temporal deep neural network. Sensors, 19(15), 3420.

17. 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.

18. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774).

19. 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(1), 119.

20. Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L., Rothchild, D., ... (2021). Carbon emissions and large neural network training. arXiv preprint arXiv:2104.10350.

21. 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).