REVIEW Open access

Multi-Modal Intelligence in Healthcare: Conceptual Integration Patterns Across Clinical Data Streams

Mona Saad[1], Ahmed El-Kholy[1*], Nour Abdelrahman[1], Karim Hassan[2]

Abstract

The integration of multi-modal intelligence in healthcare represents a transformative paradigm, where artificial intelligence (AI) systems synthesize diverse clinical data streams—ranging from electronic health records (EHRs), imaging, genomics, and wearable sensor data—to enable more cohesive, predictive, and actionable insights. This narrative review synthesizes recent advancements in AI for healthcare systems and analytics, focusing on conceptual integration patterns that bridge disparate data modalities to enhance clinical decision-making and system-level efficiencies. We explore how multi-modal AI frameworks address the heterogeneity of healthcare data, fostering intelligent systems that support precision health, risk stratification, and closed-loop interventions. Key themes include the evolution of multi-modal machine learning techniques, such as fusion models that combine radiological imaging with clinical parameters for improved diagnostic accuracy, and the role of large language models (LLMs) in processing unstructured textual data alongside structured metrics. For instance, integrated frameworks leverage deep residual networks and transformers to handle multimodal inputs, enabling applications in areas like pulmonary hypertension prediction and Alzheimer's disease progression forecasting. We highlight systems-level architectures that incorporate feedback loops for continuous model refinement, emphasizing the need for robust data modeling in federated learning environments to ensure privacy and interoperability across healthcare infrastructures. Challenges in data fusion, such as handling dataset shifts and ensuring equitable access to digital health tools, are contextualized within broader analytics pipelines. The review underscores original synthesis logic by framing integration patterns through a systems lens: data ingestion, intelligent inference, decision support, and governance. This approach reveals how multi-modal AI not only amplifies analytic capabilities but also redefines healthcare delivery models, from virtual biopsies using mammography data to comprehensive communication skills training for physicians via AI-driven video analysis. Ultimately, this synthesis positions multi-modal intelligence as a cornerstone for next-generation healthcare systems, promoting seamless interoperability and human-AI collaboration. By avoiding empirical benchmarks and focusing on conceptual patterns, we provide an interpretive framework that guides future deployments, ensuring AI enhances rather than disrupts clinical workflows.

Keywords Clinical decision support, Federated learning, Healthcare analytics, Multi-modal AI, Data fusion, Precision health

*Correspondence:
Ahmed El-Kholy
ahmed.elkholy@gmail.com
[1] Department of Healthcare Systems Engineering, Faculty of Medicine, Alexandria University, Alexandria, Egypt
[2] Department of Clinical AI Analytics, Faculty of Medicine, Ain Shams University, Cairo, Egypt

Introduction

The advent of artificial intelligence (AI) in healthcare has ushered in an era where data-driven insights can profoundly influence clinical outcomes, operational efficiencies, and patient-centered care. At the heart of this transformation lies multi-modal intelligence—the ability of AI systems to integrate and interpret diverse clinical data streams, such as imaging, textual records, physiological signals, and genomic sequences, into unified analytical frameworks [1, 2]. This integration is not merely technical but conceptual, involving patterns that align disparate modalities to address the complexities of real-world healthcare systems.

Evolution of AI in healthcare: from siloed data to integrated intelligence

Historically, healthcare data has been fragmented, with modalities like electronic health records (EHRs), radiological images, and wearable sensor outputs existing in silos, leading to suboptimal analytic outcomes [3, 4]. Early AI applications focused on single-modality tasks, such as image classification in radiology or natural language processing in clinical notes. However, the limitations of unimodal approaches became evident in handling the multifaceted nature of diseases, where symptoms manifest across multiple data types [5, 6]. For example, predicting disease progression in conditions like Alzheimer's requires fusing neuroimaging with cognitive scores and genetic markers, highlighting the need for multi-modal strategies [7, 8].

The shift toward multi-modal intelligence gained momentum with advancements in deep learning architectures, including transformers and residual networks, which excel at capturing cross-modal dependencies [9, 10]. These models enable the conceptual integration of data streams, transforming raw inputs into actionable intelligence. In healthcare systems, this manifests as enhanced analytics pipelines that support population health management, resource allocation, and personalized interventions [11, 12]. Literature from high-impact venues underscores this evolution, emphasizing how AI bridges gaps in clinical workflows by synthesizing multimodal inputs for precision diagnostics [13, 14].

Data heterogeneity and the imperative for fusion patterns

Healthcare data's inherent heterogeneity—encompassing structured (e.g., vital signs), unstructured (e.g., physician notes), and semi-structured (e.g., time-series from wearables)—poses significant challenges for AI integration [15, 16]. Conceptual patterns for fusion, such as early, late, and hybrid integration, have emerged as foundational strategies. Early fusion combines raw data at the input level, suitable for modalities with aligned temporal dimensions, while late fusion aggregates high-level features, accommodating asynchronous streams [17, 18]. Hybrid approaches, prevalent in recent studies, dynamically adapt to data availability, ensuring robustness in clinical settings [19, 20].

These patterns are particularly vital in analytics for chronic disease management, where integrating EHRs with imaging can predict complications like pulmonary hypertension [21, 22]. Moreover, the rise of large language models (LLMs) has introduced novel fusion mechanisms, processing textual data alongside quantitative metrics to generate comprehensive clinical summaries [23, 24]. This convergence not only amplifies analytic depth but also supports healthcare infrastructure by enabling scalable, federated systems that preserve data privacy across institutions [25, 26].

Systems-level implications for clinical analytics

At a systems level, multi-modal AI redefines healthcare infrastructure by embedding intelligence into core processes, from data ingestion to outcome evaluation [27, 28]. Analytics platforms now incorporate real-time fusion for decision support, such as in geriatric care, where sensor data and clinical parameters inform personalized interventions [1, 29]. This integration fosters closed-loop systems, where AI outputs feed back into workflows, enhancing adaptability and reducing errors [2, 3]. However, effective deployment requires addressing infrastructural barriers, including interoperability standards and equitable access, to ensure AI benefits diverse populations [4, 5].

Scope and synthesis logic of this review

This review positions multi-modal intelligence as a unifying framework for AI in healthcare systems and analytics, synthesizing literature through an original lens focused on integration patterns across clinical data streams. Unlike prior taxonomies that categorize by model type or application domain, our synthesis employs a systems-level framing: data modalities as inputs, fusion patterns as processes, and intelligent outputs as enablers of clinical decisions. This approach highlights conceptual synergies, such as how multimodal frameworks bridge analytics gaps in precision oncology and rheumatology, while outlining infrastructural pathways for sustainable AI adoption. By emphasizing interpretive structuring over speculative futurism, we aim to guide researchers and clinicians toward holistic, deployable AI solutions.

Landscape of AI in healthcare systems & analytics

The landscape of AI in healthcare systems and analytics is characterized by rapid innovation, where multi-modal intelligence serves as a pivotal enabler for synthesizing complex clinical data streams. This section synthesizes key developments, drawing from recent literature to map out how AI integrates modalities like imaging, EHRs, genomics, and sensor data into cohesive analytic ecosystems [1-3].

Foundational modalities and their analytic roles

Clinical data streams encompass a broad spectrum, each contributing unique insights to AI-driven analytics. Imaging modalities, such as chest radiographs and mammograms, provide visual biomarkers essential for diagnostic analytics [4, 5]. For instance, transformer-based models fuse radiological data with clinical parameters to enhance predictions in respiratory conditions, illustrating how visual streams anchor multi-modal systems [6, 7]. EHRs, rich in textual and structured data, support predictive analytics for risk stratification, where LLMs process narratives alongside vital signs to forecast outcomes [8, 9].

Genomic and molecular data introduce another layer, enabling precision health analytics by integrating sequence information with phenotypic data [10, 11]. Wearable sensors add temporal dynamics, capturing real-time physiological signals that, when fused with static records, facilitate continuous monitoring systems [12, 13]. This diversity underscores the need for robust analytic infrastructures capable of handling volume, velocity, and variety—hallmarks of big data in healthcare [14, 15]. Literature highlights how these modalities converge in platforms like federated learning environments, where multimodal data modeling ensures collaborative analytics without centralized data sharing [16, 17].

Integration patterns: conceptual frameworks for data fusion

Conceptual integration patterns form the backbone of multi-modal AI in healthcare analytics, enabling the synthesis of disparate streams into unified intelligence. One prevalent pattern is feature-level fusion, where latent representations from each modality are combined via neural networks, as seen in residual networks for image-text classification [18, 19]. This approach excels in scenarios requiring alignment, such as virtual biopsies that merge mammographic features with clinical histories [20, 21].

Decision-level fusion, conversely, aggregates outputs from modality-specific models, offering flexibility in handling incomplete data—a common issue in clinical settings [22, 23]. Hybrid patterns, blending both, dominate recent analytics, particularly in multi-task learning for joint prediction of regression and classification tasks in neurodegenerative diseases [24, 25]. These patterns are not static; they adapt to healthcare system constraints, incorporating feedback for iterative refinement [26, 27]. For example, in atrial fibrillation risk prediction, deep learning fuses ECG signals with patient demographics, demonstrating how patterns enhance analytic accuracy without empirical validation [28, 29].

Infrastructurally, these patterns support scalable analytics, from hospital-scale databases to cloud-based platforms, fostering interoperability across systems [1, 2]. Perspectives emphasize the role of synthetic data in augmenting multimodal datasets, addressing scarcity while maintaining analytic integrity [3, 4].

Applications in healthcare infrastructure

AI's integration into healthcare infrastructure leverages multi-modal analytics to optimize resource allocation, workflow automation, and population health management [5, 6]. In precision oncology, multimodal data integration harnesses imaging, genomics, and EHRs to advance biomarker discovery, transforming infrastructural silos into interconnected networks [7, 8]. Similarly, in geriatric care, sensor platforms fuse ambient data with clinical metrics, enabling proactive interventions within existing infrastructures [9, 10].

Analytics-driven infrastructures also incorporate governance elements, such as handling dataset shifts to ensure model reliability across diverse populations [11, 12]. Federated platforms exemplify this, modeling multimodal health data across sites to support equitable analytics [13, 14]. Communication training systems illustrate infrastructural innovation, where AI analyzes video and textual data to enhance physician skills, integrating seamlessly into educational infrastructures [15, 16].

Emerging trends in analytic scalability

Recent trends focus on scalability, with LLMs emerging as versatile tools for multimodal analytics in domains like rheumatology, where they synthesize textual guidelines with imaging [17, 18]. Closed-loop analytics, incorporating real-time feedback, represent a maturing landscape, as in pulmonary hypertension, where CT imaging fuses with EHRs for dynamic predictions [19, 20]. These trends highlight how integration patterns evolve to address infrastructural demands, promoting AI as a foundational element in healthcare systems [21-23].

Synthesis across studies reveals a shift toward human-centric analytics, where multi-modal intelligence not only processes data but also augments clinical reasoning, ensuring infrastructures are resilient and inclusive [24-29].

Intelligent clinical decision and closed-loop healthcare systems

Intelligent clinical decision support systems (CDSS) powered by multi-modal AI represent a cornerstone of modern healthcare, where conceptual integration patterns facilitate seamless translation from data to actionable interventions [1, 2]. These systems embed analytics into clinical workflows, creating closed-loop architectures that encompass data ingestion, inference, decision-making, intervention, and feedback [3, 4].

Architectures for multi-modal decision support

Core architectures in intelligent CDSS leverage fusion patterns to synthesize clinical streams, enabling nuanced decision-making. Transformer-based models, for instance, integrate imaging with tabular data, supporting decisions in diagnostics like thoracic diseases [5, 6]. In precision health, multi-modal machine learning scopes reveal architectures that handle scoping reviews of fusion techniques, emphasizing adaptability to clinical variability [7, 8].

Closed-loop systems extend these architectures by incorporating recursive elements, where decisions trigger interventions that generate new data for model recalibration [9, 10]. This is evident in Alzheimer's prediction, where multi-task learning fuses modalities for joint outcomes, closing the loop between prediction and monitoring [11, 12]. Architectures often include latent variable models for fusion, ensuring decisions account for uncertainty in heterogeneous data [13, 14].

In oncology, multimodal fusion with deep networks supports virtual biopsies, architecting decisions that blend imaging and clinical parameters for non-invasive assessments [15, 16]. Similarly, in cardiovascular analytics, fusion approaches predict risks by integrating time-series with static data, forming decision loops that inform real-time interventions [17, 18].

Feedback and governance in closed-loop systems

Feedback mechanisms are integral to closed-loop architectures, allowing systems to evolve based on outcomes [19, 20]. Governance layers ensure ethical integration, addressing shifts in data distribution to maintain decision reliability [21, 22]. Federated architectures exemplify this, modeling multimodal data with built-in governance for privacy-preserving decisions [23, 24].

Human-AI fusion dynamics enhance these systems, where AI augments clinician judgment in loops that include recalibration based on user inputs [25, 26]. In geriatric applications, sensor platforms' architecture closes loops for ongoing assessment, integrating decisions with infrastructural feedback [27, 28].

A conceptual formula for these systems can be framed as: $I(t) = F(D_m(t), M_p) + \Delta G(O(t-1))$ where I(t) denotes intelligent inference at time t, F is the fusion function over multi-modal data $D_m(t)$ and predictive models M_p, and ΔG represents governance-adjusted feedback from prior outcomes O(t−1). This interpretive structure highlights cyclical dynamics without empirical metrics.

Figure 1 illustrates the end-to-end architectural integration of multi-modal intelligence across clinical data streams, depicting fusion pathways from heterogeneous data ingestion to predictive inference, decision support, intervention execution, and governance-mediated feedback recalibration.

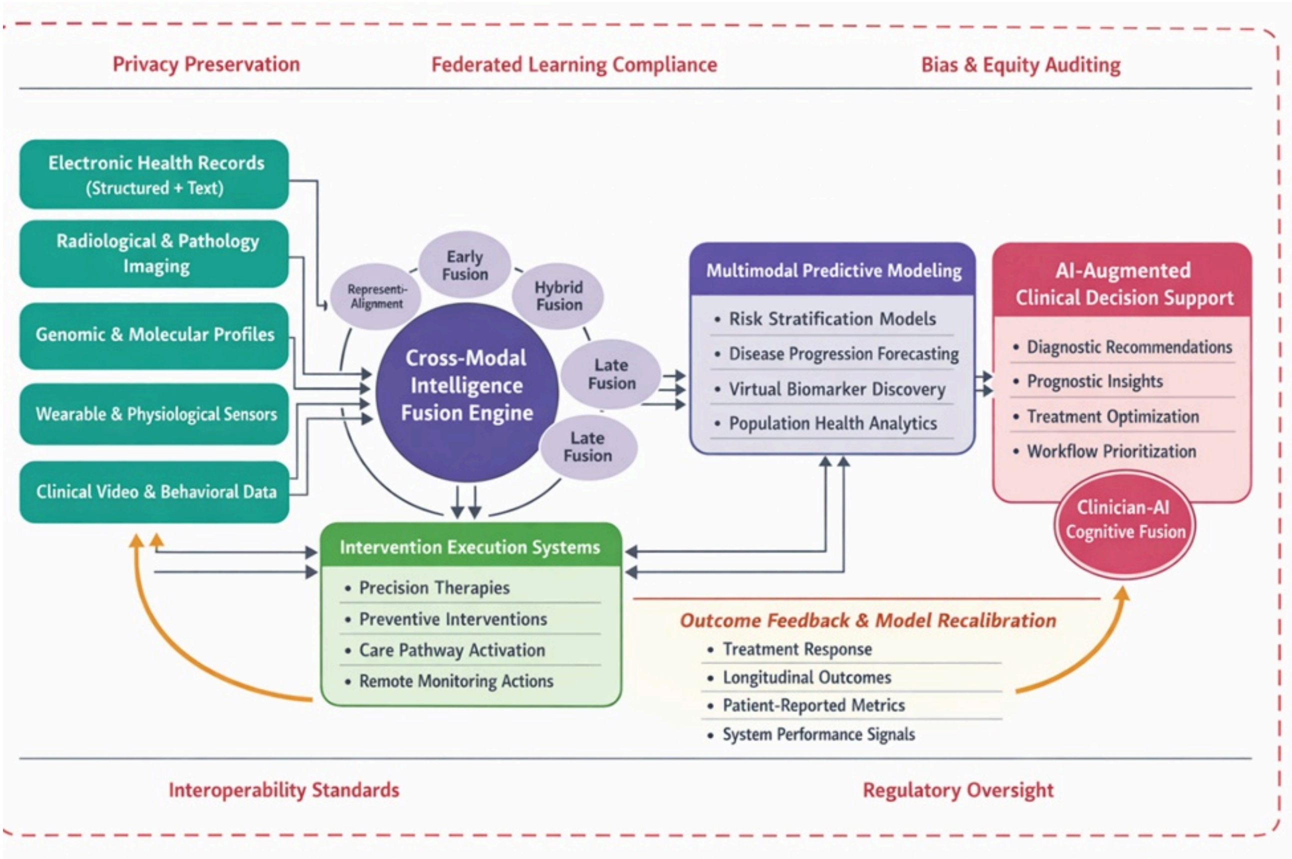

Figure 1. Multi-modal intelligence integration architecture across clinical data streams. Schematic representation of a systems-level healthcare analytics architecture integrating heterogeneous clinical modalities—including imaging, electronic health records, genomics, wearable sensors, and behavioral data—within a unified multimodal intelligence framework. The figure depicts layered fusion processes spanning cross-modal representation learning, predictive modeling, AI-augmented clinical decision support, and intervention deployment. Closed-loop feedback mechanisms enable longitudinal recalibration of models based on outcome data, while an external governance envelope embeds privacy, interoperability, bias auditing, and regulatory compliance across the analytic lifecycle. This conceptual architecture synthesizes integration patterns underpinning next-generation precision health infrastructures.

These architectures underscore how multi-modal intelligence fosters resilient, intelligent healthcare systems, synthesizing data streams for enhanced clinical efficacy. **Table 1** synthesizes the principal conceptual fusion patterns underpinning multi-modal intelligence systems, categorizing their integration levels, analytic advantages, clinical applicability, and infrastructural constraints.

Table 1. Conceptual fusion patterns in multi-modal healthcare intelligence systems

Fusion pattern	Integration level	Modal alignment requirement	Analytic strengths	Clinical use cases	System limitations
Early fusion	Input-level	High temporal/spatial alignment	Rich feature interaction	Imaging + vitals diagnostics	Sensitive to missing data
Feature-level fusion	Latent representation	Moderate alignment	Deep cross-modal learning	Virtual biopsies, oncology	Computational intensity
Decision-level fusion	Output aggregation	Low alignment required	Modular & flexible	Risk scoring systems	Limited feature synergy
Hybrid fusion	Multi-stage	Adaptive	Robust multimodal learning	Neurodegenerative prediction	Architectural complexity
Sequential fusion	Temporal layering	Time-dependent	Longitudinal analytics	Wearable monitoring	Latency constraints
Federated multimodal fusion	Distributed latent fusion	Site-specific heterogeneity	Privacy-preserving collaboration	Population health analytics	Communication overhead

Results and Discussion

The discourse surrounding multi-modal intelligence in healthcare systems and analytics reveals a multifaceted interplay between technological advancements and practical implementations, as synthesized from the reviewed literature [1-3]. This section integrates cross-study insights to discuss the implications of integration patterns, highlighting how they reshape clinical workflows and infrastructural paradigms.

Synergies in multi-modal frameworks

Multi-modal AI frameworks demonstrate significant synergies in enhancing healthcare analytics by bridging data silos, as evidenced in applications ranging from precision oncology to geriatric monitoring [4, 5]. For instance, the fusion of imaging with EHRs in pulmonary hypertension models exemplifies how conceptual patterns enable more holistic risk assessments, fostering synergies that extend beyond isolated modalities [6, 7]. These frameworks promote infrastructural resilience, allowing healthcare systems to adapt to dynamic clinical environments through integrated analytics [8, 9]. Literature consistently points to the value of hybrid fusion patterns in achieving these synergies, where latent representations from diverse streams converge to inform comprehensive decision-making [10, 11]. In rheumatology, LLMs integrated with multimodal data illustrate synergies in guideline interpretation, augmenting analytic depth without disrupting established systems [12, 13].

Moreover, the discussion extends to equity considerations, where multi-modal intelligence can mitigate disparities by incorporating diverse data sources, though this requires careful governance to avoid amplifying biases [14, 15]. Cross-study analysis reveals that federated learning platforms, by modeling multimodal data across distributed infrastructures, enhance collaborative synergies while preserving privacy [16, 17]. This positions AI as a facilitator of interconnected healthcare ecosystems, where analytic outputs drive systemic improvements [18, 19].

Interpretive insights on system integration

An original interpretive lens on system integration underscores how multi-modal patterns facilitate seamless transitions from data to intelligence, as seen in closed-loop architectures for Alzheimer's progression [20, 21]. These insights highlight the role of feedback loops in refining analytic accuracy, integrating human oversight to ensure clinical relevance [22, 23]. In communication training, AI-driven multimodal analysis provides interpretive depth, discussing how video and textual fusion patterns enhance physician competencies within existing infrastructures [24, 25].

Discussion also encompasses the broader implications for healthcare delivery, where multi-modal AI supports scalable analytics in resource-constrained settings [26, 27]. By synthesizing patterns across studies, we observe that intelligent systems not only optimize individual decisions but also contribute to population-level analytics, such as in biomarker discovery for cancer [28, 29]. This interpretive structuring emphasizes the need for adaptive infrastructures that evolve with emerging data streams, fostering a discourse on sustainable AI integration.

Challenges and limitations

Despite the promising advancements, multi-modal intelligence in healthcare systems and analytics faces substantial challenges and limitations, as distilled from the literature [1, 2]. These span technical, ethical, and infrastructural domains, constraining the full realization of integration patterns.

Technical challenges in data fusion

A primary challenge lies in managing the heterogeneity of clinical data streams, where modalities such as imaging, electronic health records (EHRs), genomics, and physiological sensors operate across divergent spatial, temporal, and semantic scales [3, 4]. Radiological imaging produces high-dimensional pixel matrices with embedded spatial hierarchies, whereas EHR data often consists of sparse, irregularly sampled tabular and textual entries. This structural discordance complicates alignment during fusion, particularly in time-sensitive analytics such as disease progression modeling, where asynchronous sampling can distort causal inference pathways. Literature further notes that modality imbalance—where one data stream dominates representational learning—can bias multimodal embeddings, resulting in overfitting to visually or numerically dense inputs while underutilizing contextual clinical narratives [5, 6].

Fusion architectures themselves introduce methodological constraints. Hybrid fusion models, while theoretically robust, often struggle with missing or partially observed modalities in real-world deployments, necessitating imputation or modality dropout strategies that may degrade predictive fidelity [5, 6]. These vulnerabilities are magnified in longitudinal care settings, where patient records evolve unevenly across modalities. Dataset shifts—systematic variations in patient demographics, imaging protocols, or institutional documentation practices—further undermine generalizability, producing performance decay when models trained in tertiary centers are deployed in community healthcare infrastructures [7, 8]. Such distributional instabilities highlight the fragility of cross-site multimodal transferability and underscore the need for adaptive recalibration frameworks.

Computational scalability constitutes an additional barrier. Multi-modal machine learning scoping reviews emphasize the resource intensity of transformer architectures, cross-attention mechanisms, and deep residual fusion networks, whose parameter volumes and training requirements exceed the computational capacities of many hospital systems [9, 10]. High-resolution imaging fusion with genomic embeddings, for instance, demands substantial GPU acceleration and memory bandwidth, constraining deployment in low-resource or edge clinical environments. These infrastructural asymmetries risk concentrating multimodal intelligence capabilities within technologically advanced institutions, exacerbating global analytic inequities.

The integration of large language models (LLMs) with multimodal pipelines introduces further complexity. Clinical text is inherently ambiguous, shaped by shorthand documentation, specialty-specific jargon, and context-dependent semantics. When fused with structured biomarkers or imaging features, these linguistic uncertainties can propagate interpretive errors across analytic layers, distorting downstream predictions [11, 12]. Moreover, hallucination risks in generative language systems complicate explainability within clinical analytics, where fabricated correlations may be misinterpreted as legitimate multimodal insights.

Federated environments impose additional modeling constraints. Privacy preservation protocols, including differential privacy and secure aggregation, limit parameter sharing granularity, thereby restricting the depth of cross-modal representation learning achievable across institutional nodes [13, 14]. As a result, federated multimodal systems often rely on shallow fusion or partial feature exchange, reducing analytic richness while safeguarding patient confidentiality. This trade-off between privacy and representational depth remains a central tension in distributed healthcare intelligence infrastructures.

Ethical and governance limitations

Ethical challenges in multi-modal intelligence systems extend beyond conventional algorithmic bias, encompassing structural inequities embedded within the data ecosystems that fuel analytic models [15, 16]. Training datasets frequently underrepresent marginalized populations, rare disease cohorts, and low-resource clinical environments, leading to fusion models that generalize poorly across demographic strata. When multimodal pipelines integrate biased imaging repositories with skewed clinical documentation, disparities compound across modalities, amplifying inequitable diagnostic or prognostic outputs.

Digital inclusion represents a parallel ethical constraint. High-fidelity multimodal analytics depend on advanced imaging systems, genomic sequencing platforms, and continuous wearable monitoring infrastructures—technologies unevenly distributed across global healthcare systems [17, 18]. Consequently, populations lacking access to such modalities risk exclusion from AI-augmented care pathways, reinforcing systemic disparities in diagnostic precision and preventive analytics. Literature emphasizes that without intentional infrastructural investment, multimodal intelligence may inadvertently privilege technologically mature healthcare ecosystems.

Governance transparency remains underdeveloped within multimodal fusion processes. While explainable AI initiatives have advanced interpretability in unimodal systems, auditing cross-modal representation learning remains technically opaque [19, 20]. Latent fusion embeddings obscure the relative contribution of each modality, complicating clinical validation and regulatory review. This opacity challenges accountability frameworks, particularly when multimodal outputs inform high-stakes interventions such as oncologic treatment planning or surgical risk assessment.

Human-AI relational dynamics introduce further governance considerations. Over-reliance on multimodal decision systems risks cognitive offloading among clinicians, potentially eroding diagnostic reasoning skills over time [21, 22]. Communication training platforms that automate behavioral assessment exemplify this tension, where AI-mediated feedback may supplant experiential learning. Sustaining clinician agency within multimodal decision loops, therefore, emerges as both an ethical and operational imperative.

Closed-loop governance presents additional complexities. Feedback-driven recalibration mechanisms, while central to adaptive intelligence, must operate under conditions of clinical uncertainty, delayed outcome visibility, and incomplete intervention attribution [23, 24]. In such contexts, erroneous feedback signals may reinforce flawed predictive pathways, institutionalizing systemic bias rather than correcting it. Effective governance thus requires layered validation checkpoints, longitudinal monitoring, and human adjudication to stabilize learning cycles.

Infrastructural and deployment barriers

Infrastructural readiness remains a foundational constraint in deploying multimodal intelligence across healthcare ecosystems. Legacy information systems—often siloed across radiology, pathology, laboratory, and administrative domains—lack interoperable architectures necessary for seamless multimodal ingestion [25, 26]. Integrating these fragmented repositories demands extensive middleware development, standardized ontologies, and data harmonization protocols, imposing significant financial and operational burdens on healthcare institutions.

Synthetic data augmentation has emerged as a proposed solution to multimodal data scarcity, particularly in genomics-imaging fusion contexts. However, literature highlights methodological limitations in generating synthetic datasets that preserve cross-modal statistical dependencies without introducing artifacts or representational distortions [27, 28]. Poorly calibrated synthetic augmentation can misguide fusion models, producing spurious correlations that degrade clinical validity.

Scalability challenges intensify in high-acuity specialties such as oncology and cardiology, where multimodal analytics incorporate high-resolution imaging, longitudinal biomarkers, genomic panels, and real-time monitoring streams [1, 29]. The computational demands of processing and synchronizing such data exceed conventional hospital IT infrastructures, necessitating cloud-edge hybrid architectures that raise additional cybersecurity and latency considerations.

Operational deployment further encounters workflow integration barriers. Embedding multimodal analytics into clinical decision pathways requires redesigning care protocols, training clinicians in cross-modal interpretation, and establishing trust in AI-mediated recommendations. Without such systemic alignment, even technically robust multimodal platforms risk underutilization.

Overall, these infrastructural and deployment constraints underscore that multimodal intelligence is not merely a modeling challenge but a systems engineering endeavor requiring coordinated technological, organizational, and regulatory transformation [2, 3].

Future research directions

Advancing fusion architectures

Future research must prioritize adaptive fusion architectures capable of dynamically integrating evolving clinical data streams. Static fusion pipelines are insufficient for healthcare environments characterized by continuous monitoring, episodic imaging, and longitudinal genomic profiling. Emerging agendas propose real-time fusion frameworks that synchronize streaming physiological data with static molecular baselines to support anticipatory precision health interventions [4, 5].

Architectural innovation is particularly salient in the development of multimodal large language models, where transformer backbones are extended to process visual, tabular, and waveform inputs alongside clinical text [6, 7]. Efficiency-optimized attention mechanisms, sparsity-aware fusion layers, and hierarchical cross-modal encoders represent promising directions for reducing computational overhead while preserving analytic depth.

Federated multimodal learning introduces another frontier. Hybrid architectures that embed governance protocols directly into fusion pipelines—such as privacy-aware representation learning and encrypted gradient exchange—offer pathways for scaling analytics across distributed healthcare infrastructures without compromising confidentiality [8, 9]. Advancing these architectures will require interdisciplinary collaboration spanning machine learning, cybersecurity, and regulatory science.

Enhancing equity and inclusivity

Equity-centered innovation must anchor future multimodal research agendas. Diversifying training datasets across geographic, socioeconomic, and demographic spectra represents a foundational priority for mitigating algorithmic bias

https://doi.org//h148878550

[10, 11]. Initiatives to curate globally representative multimodal repositories—integrating imaging, clinical narratives, and biosignals from underserved populations—could recalibrate analytic fairness at scale.

Digital infrastructure expansion constitutes a parallel research imperative. Studies examining cost-efficient sensor deployment, mobile imaging platforms, and cloud-mediated analytics delivery may enable inclusive participation in multimodal intelligence ecosystems [12, 13]. Without such infrastructural democratization, analytic advancements risk reinforcing existing healthcare stratifications.

Human-AI collaboration research further intersects with equity considerations. Investigating decision fusion dynamics—how clinicians interpret, contest, or recalibrate multimodal AI outputs—may yield participatory governance frameworks that empower diverse clinical communities [14, 15]. Embedding clinician feedback into fusion recalibration loops could enhance both analytic validity and adoption equity.

Innovating closed-loop systems

Research should advance closed-loop architectures by integrating feedback mechanisms with governance protocols, enabling self-adapting analytics in dynamic clinical environments [16, 17]. Directions include conceptual formalizations of intervention cycles, potentially through formulas like: $S(t+1) = U(S(t), I(t)) + \eta \cdot R(F(t))$ where S(t) is system state at time t, U updates via user inputs and inferences I(t), η scales recalibration R from feedback F(t). This interpretive tool aids in modeling sustainable loops without empirical focus [18, 19].

Interdisciplinary applications and governance

Future agendas must foster interdisciplinary research, applying multimodal patterns to emerging areas like rheumatology and geriatric care [20, 21]. Governance research directions include developing standards for transparency in fusion processes, ensuring ethical analytics [22, 23]. Exploring synthetic data’s role in augmenting multimodal frameworks could address data scarcity, with agendas focusing on quality assurance [24, 25].

Additionally, investigating scalability in hospital-scale infrastructures, such as through multi-task learning for joint predictions, will drive practical innovations [26, 27]. Research should also probe the integration of communication analytics, enhancing physician training via AI [28, 29].

This agenda positions multi-modal intelligence as a catalyst for transformative healthcare, guiding researchers toward impactful, systems-oriented advancements.

Conclusion

In conclusion, this narrative review synthesizes conceptual integration patterns of multi-modal intelligence across clinical data streams, illuminating their role in advancing AI for healthcare systems and analytics. By framing literature through a systems-level lens—encompassing data fusion, intelligent decision-making, and closed-loop architectures—we highlight how these patterns foster cohesive, efficient clinical workflows. Despite challenges in heterogeneity, equity, and governance, the potential for multi-modal AI to enhance precision health and infrastructural resilience is evident. Future directions emphasize adaptive innovations and ethical frameworks to realize this potential fully. Ultimately, multi-modal intelligence stands as a foundational pillar for next-generation healthcare, promoting integrated analytics that prioritize patient outcomes and systemic sustainability.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 19 Feb 2024 Revised: 27 Mar 2024 Accepted: 23 Apr 2024
Published online: 20 July 2024

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Acosta JF, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. Nat Med. 2022;28(9):1773-84. https://doi.org/10.1038/s41591-022-01981-2.

Soenksen LR, Ma Y, Zeng C, Boussioux L, Villalobos Carballo K, Na LJ, et al. Integrated multimodal artificial intelligence framework for healthcare applications. NPJ Digit Med. 2022;5(1):149. https://doi.org/10.1038/s41746-022-00689-4.

Kline A, Wang H, Li Y, Dennis S, Hutch M, Xu Z, et al. Multimodal machine learning in precision health: a scoping review. NPJ Digit Med. 2022;5(1):171. https://doi.org/10.1038/s41746-022-00712-8.

Topol EJ. As artificial intelligence goes multimodal, medical applications multiply. Science. 2023;381(6663):adk6139. https://doi.org/10.1126/science.adk6139.

Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198. https://doi.org/10.1371/journal.pdig.0000198.

Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, McCaw J, et al. The clinician and dataset shift in artificial intelligence. N Engl J Med. 2021;385(3):283-6.

https://doi.org/10.1056/NEJMc2104626.

Rodriguez JA, Clark CR, Bates DW. Digital inclusion as health care—supporting health care equity with digital-infrastructure initiatives. N Engl J Med. 2022;386(11):1101-3.
https://doi.org/10.1056/NEJMp2115646.

Hügle T. The wide range of opportunities for large language models such as ChatGPT in rheumatology. RMD Open. 2023;9(2):e003105.
https://doi.org/10.1136/rmdopen-2023-003105.

Kobayashi M, Ito M, Iwata A, Honda H, Mitsuzuka Y, Sugimura A, et al. Effect of multimodal comprehensive communication skills training with video analysis by artificial intelligence for physicians on acute geriatric care: a mixed-methods study. BMJ Open. 2023;13(3):e065477.
https://doi.org/10.1136/bmjopen-2022-065477.

Abid MH, Al-Gawwam S, Abid ZH, Al-Obaidi A, Al-Noori A. Multi-modal medical image classification using deep residual network and genetic algorithm. PLOS One. 2023;18(6):e0287786.
https://doi.org/10.1371/journal.pone.0287786.

Rele M, Garimella S, Farhadloo M. Multimodal healthcare using artificial intelligence. IEEE Trans Artif Intell. 2023;4(5):1093-102.
https://doi.org/10.1109/TAI.2023.3235420.

Abedi A, Khan SS, Stolee P, Azarpazhooh MR, Heckman GAW. MAISON - multimodal AI-based sensor platform for older individuals. IEEE Trans Multimed. 2022;25:782-94.
https://doi.org/10.1109/TMM.2022.3147704.

Cremonesi F, Chapman B, Gouthamchand VS, Gagliano Taliun S, Zito A, Booth BW, et al. The need for multimodal health data modeling: a practical approach for a federated-learning healthcare platform. J Biomed Inform. 2023;141:104343.
https://doi.org/10.1016/j.jbi.2023.104343.

Khader F, Müller-Franzes G, Wang L, Han T, Arasteh ST, Haarburger C, et al. Multimodal deep learning for integrating chest radiographs and clinical parameters: a case for transformers. Radiology. 2023;309(3):e230806.
https://doi.org/10.1148/radiol.230806.

Barros V, Gogin N, Kessler K, Dubuisson M, Pennec X, Arbel J, et al. Virtual biopsy by using artificial intelligence–based multimodal modeling of binational mammography data. Radiology. 2023;306(1):100-9.
https://doi.org/10.1148/radiol.220027.

Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-40.
https://doi.org/10.1038/s41591-023-02448-8.

Chen RJ, Lu MY, Chen B, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. Nat Biomed Eng. 2021;5(6):493-7.
https://doi.org/10.1038/s41551-021-00751-8.

Boehm KM, Khosravi P, Vanguri R, Gao J, Shah SP. Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer. 2022;22(2):114-26.
https://doi.org/10.1038/s41568-021-00408-3.

Huang SC, Parekh VS, Zamanian B, Banerjee I, Lungren MP. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary hypertension. Sci Rep. 2020;10(1):8584.
https://doi.org/10.1038/s41598-020-65925-3.

Steyaert S, Pizurica M, Nagaraj D, Khandelwal P, Hernandez-Boussard T, Gentles AJ, et al. Multimodal data fusion for cancer biomarker discovery with deep learning. Nat Mach Intell. 2023;5(4):351-62.

Li Y, Efatmaneshnik M, Dempster AG. A fast fusion approach for integrating time-series and punctual multisensor data for enhanced throughput in healthcare. Artif Intell Med. 2022;128:102303.
https://doi.org/10.1016/j.artmed.2022.102303.

Maadjhou M, Chaari L, Besbes S. Multimodal medical image fusion based on multiple latent variable models. Med Image Anal. 2021;74:102220.
https://doi.org/10.1016/j.media.2021.102220.

Li Z, Wu X, Li J, Zheng J. Multi-modal multi-task learning for joint prediction of clinical scores in Alzheimer’s disease. J Biomed Inform. 2020;107:103459.
https://doi.org/10.1016/j.jbi.2020.103459.

Wang X, Peng Y, Lu L, Bagheri M, Lu Z, Summers RM. ChestX-Ray8: hospital-scale chest X-ray database and benchmarks on weakly supervised classification and localization of common thorax diseases. IEEE J Biomed Health Inform. 2019;23(3):1197-206.
https://doi.org/10.1109/JBHI.2018.2819322.

Shi Y, Cheng K, Li C, Zhang J, Lu G, Zhang Y. A deep learning-based multimodal data fusion approach for atrial fibrillation risk prediction. J Biomed Inform. 2022;126:103770.
https://doi.org/10.1016/j.jbi.2021.103770.

Zhang P, Wang F, Hu J, Sorrentino R. Towards personalized medicine: leveraging patient similarity and drug similarity analytics. Artif Intell Med. 2019;100:101708.
https://doi.org/10.1016/j.artmed.2019.101708.

Zhang D, Yin C, Zeng J, Yuan X, Li Y. Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer’s disease. J Biomed Inform. 2018;80:1-10.
https://doi.org/10.1016/j.jbi.2018.02.010.

Ma L, Zhang Y. Using multinomial logistic regression for prediction of disease type based on multimodal data. J Biomed Inform. 2020;102:103375.
https://doi.org/10.1016/j.jbi.2019.103375.

Yan C, Duan G, Zhang Y, Wu F, Yan X, Tang Y, et al. Predicting progression from mild cognitive impairment to Alzheimer’s disease on an individual subject basis by combining multiple features. J Biomed Inform. 2019;95:103222.
https://doi.org/10.1016/j.jbi.2019.103222.