Review

Multimodal Generative AI in Diagnostics: Bridging Medical Imaging and Clinical Reasoning

Morteza Maleki * and SeyedAli Ghahari

Posted Date: 1 September 2025

doi: 10.20944/preprints202508.1425.v2

Keywords: multimodal generative AI; medical diagnostics; clinical reasoning; medical imaging; AI ethics; regulatory compliance; synthetic data; human-AI collaboration; precision medicine

Review

Multimodal Generative AI in Diagnostics: Bridging Medical Imaging and Clinical Reasoning

Morteza Maleki [1,*] **and SeyedAli Ghahari** [2]

[1] Adjunct Researcher, Emory University, Winship Cancer Institute, Georgia, USA
[2] Researcher, Institute for Advanced Construction and Smart Infrastructure Solutions of America
[*] Correspondence: mmalek3@emory.edu

Abstract

Multimodal generative artificial intelligence (AI) has emerged as a transformative technology in clinical diagnostics, integrating diverse data sources—medical imaging, genomic profiles, clinical narratives, and electronic health records—to significantly enhance diagnostic accuracy, clinical decision-making, and personalized patient care. This review systematically explores the landscape of multimodal AI across key medical specialties, including radiology, pathology, dermatology, ophthalmology, neurology, and oncology, highlighting recent methodological advancements, performance evaluations, and practical clinical implementations. Technical strategies such as tool-use, grafting, and unified multimodal architectures are critically assessed, identifying their strengths and limitations concerning clinical applicability, interpretability, and computational efficiency. Synthetic multimodal data generation methodologies—Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), diffusion models, and large language models (LLMs)—are evaluated for addressing data scarcity in rare disease research, enhancing international collaboration, and mitigating privacy concerns. Additionally, this review addresses pivotal ethical, regulatory, and liability challenges, emphasizing fairness, transparency, and accountability in AI-driven clinical diagnostics. Strategic priorities for future research are identified, including rigorous prospective clinical validation, development of standardized multimodal datasets, enhanced model interpretability, and robust regulatory frameworks. Ultimately, realizing the transformative potential of multimodal generative AI in clinical practice will require interdisciplinary collaboration among clinicians, researchers, ethicists, regulators, and patient advocacy groups, ensuring these powerful tools effectively augment human expertise, improve healthcare delivery, and advance precision medicine.

Keywords: multimodal generative AI; medical diagnostics; clinical reasoning; medical imaging; AI ethics; regulatory compliance; synthetic data; human-AI collaboration; precision medicine

1. Introduction

Artificial intelligence (AI) has rapidly become a foundational technology in healthcare, driving advances across diagnostic precision, therapeutic planning, clinical research, and public health interventions. Recent work illustrates its broad applicability: predictive modeling for patient readmission, protocol automation in clinical trials, and disease identification through advanced signal processing all highlight the growing integration of AI into modern healthcare practice [1–9]. Beyond direct clinical use, AI-driven analytics have provided novel insights into social determinants of health, health equity, vaccine hesitancy, and healthcare accessibility, with implications for both health policy and pandemic preparedness [10–14]. Economic modeling that links health outcomes with market and financial indicators has further expanded the role of AI in assessing system sustainability and resilience during crises [15]. Building on these foundations, **multimodal generative AI** has emerged as the next frontier in diagnostic medicine. By jointly analyzing diverse data sources—such as radiology, pathology, clinical narratives, genomics, and physiologic signals—these models promise significant enhancements in diagnostic accuracy, interpretability, and personalized care. The overall clinical pipeline for multimodal diagnostic AI, from data ingestion and fusion to deployment and monitoring, is summarized in Figure 1. A broader perspective on the healthcare AI lifecycle, spanning data curation through governance and regulation, is illustrated in Figure 2. Together, these frameworks underscore how multimodal AI approaches are positioned to transform medical diagnostics and enable safer, more effective integration into clinical workflows.

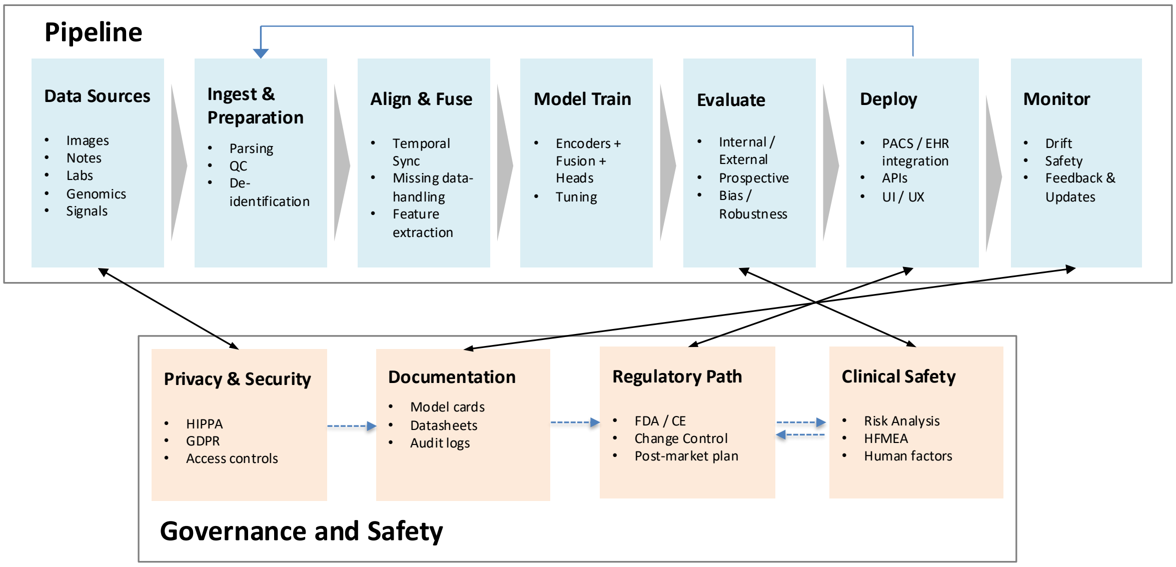


Figure 1. Overview of a clinical multimodal AI pipeline (ingest, preprocess, train/validate, deploy, monitor) with a parallel governance lane spanning reporting/validation (CONSORT-AI, SPIRIT-AI, DECIDE-AI, TRIPOD+AI) and post-deployment quality improvement/monitoring, aligned with clinical practice and safety requirements [16–22].

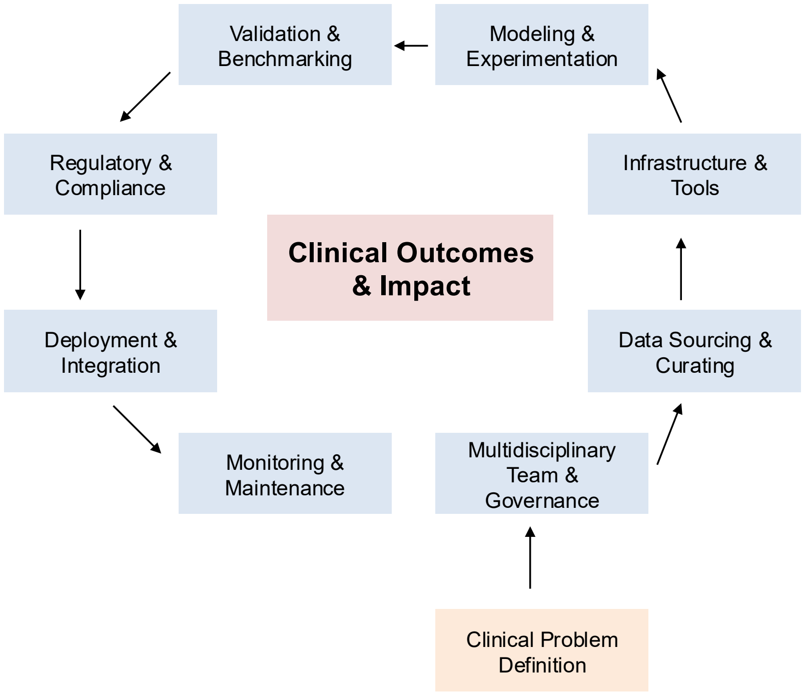

Validation & Benchmarking
Modeling & Experimentation
Regulatory & Compliance
Infrastructure & Tools
Clinical Outcomes & Impact
Deployment & Integration
Data Sourcing & Curating
Monitoring & Maintenance
Multidisciplinary Team & Governance
Clinical Problem Definition

Figure 2. Healthcare AI lifecycle spanning problem definition, data curation, model development, validation, clinical evaluation, deployment, and continuous post-market monitoring, highlighting alignment with contemporary reporting/validation frameworks [17–23].

The integration of multimodal generative artificial intelligence (AI), which combines medical imaging data with textual and clinical information, is rapidly reshaping medical diagnostics by overcoming the limitations of unimodal approaches. Medical images such as radiographs, CT scans, MRIs, and histopathology slides provide crucial anatomical and physiological insights but often lack the interpretative context supplied by patient histories, laboratory results, and diagnostic reports. By jointly analyzing these heterogeneous data streams, multimodal AI enables more comprehensive diagnostic reasoning and has demonstrated significant improvements in diagnostic accuracy, workflow efficiency, and patient outcomes [24,25]. Notable benefits are observed across medical specialties. In oncology, for instance, integration of PET–MRI or PET–CT imaging with clinical records enhances tumor characterization and informs treatment planning [26]. Similarly, in pathology, multimodal systems combining histopathological images with genomic data improve cancer grading, prognosis prediction, and recommendations for personalized therapy [27,28]. These integrated approaches provide clinicians with richer, contextually grounded insights than unimodal systems, particularly in complex diagnostic scenarios. Despite these advances, substantial challenges remain. The integration of heterogeneous multimodal data is technically complex, resource-intensive, and frequently hindered by limitations in data availability, annotation quality, and privacy protection [29]. Sophisticated modeling architectures further increase computational demands, complicating clinical deployment in resource-constrained settings. Ensuring transparency and interpretability of multimodal models is also critical for fostering clinician trust, yet remains difficult given the inherent complexity of integrating diverse modalities [24,30].

A comparative overview of clinical applications across specialties is summarized in Table 1, highlighting the spectrum of integrated data types, clinical benefits, and persistent challenges.

Table 1. Clinical applications of multimodal AI by specialty, including inputs, benefits, and challenges.

Specialty	Clinical Use-cases / Applications	Integrated Data Types (Inputs)	Clinical Benefits	Challenges
Radiology	Tumor characterization; disease classification; automated report generation [25,31]; lung nodule detection [25,32]	PET, MRI, CT, radiographs, EHR, textual reports, clinical data	Earlier cancer detection; improved diagnostic accuracy; streamlined workflow	Data integration complexity; interpretability; large-scale validation; computational demands
Pathology	Cancer grading; prognosis prediction; rare disease diagnostics [27,28,33]	Histopathology images, genomic data, clinical notes	Precision diagnostics; personalized therapy/medicine; improved prognosis prediction	Model transparency; ethical/privacy concerns; standardized data quality; data standardization
Dermatology	Skin cancer detection; lesion classification; patient education [34,35]	Clinical images, dermoscopy, patient metadata	Enhanced diagnostic accuracy; reduced unnecessary biopsies	Data bias; workflow integration; interpretability
Ophthalmology	Diabetic retinopathy detection; glaucoma monitoring; disease progression analysis [36,37]	Fundus images, OCT scans, clinical data	Earlier disease detection; accurate monitoring	Data standardization; clinical integration; ethical implications
Neurology	Alzheimer's diagnosis [27]	MRI, PET, clinical notes	Early detection; accurate progression tracking	Handling incomplete modalities; integration complexity
Oncology	Tumor characterization [26,38]	PET–MRI, PET–CT, genomic data	Improved functional assessments; tailored treatments	Regulatory hurdles; data privacy

Beyond specialty-specific advances, multimodal AI closely mirrors physician reasoning processes by synthesizing imaging findings, laboratory results, and clinical narratives into unified diagnostic hypotheses. This alignment enhances clinical workflows, improves decision-making, and uncovers subtle diagnostic cues often missed by unimodal analyses [39,40]. For example, radiology systems that combine imaging with electronic health records have been shown to streamline diagnostic reporting, reduce errors, and improve efficiency [26,41]. However, clinical integration faces persistent obstacles. Clinician acceptance depends heavily on model interpretability, usability, and demonstrable impact, yet these attributes are often undermined by model complexity [31,40]. Ethical challenges, including privacy risks, security vulnerabilities, and the propagation of algorithmic biases, further complicate deployment and acceptance [29,42]. Addressing these barriers will require the development of interpretable and standardized modeling approaches, rigorous validation frameworks, and robust governance. Promising solutions include federated learning for privacy-preserving collaboration, explainable AI methods for interpretability, and user-centered interface design to enhance clinical usability [31,39,43].

In summary, multimodal generative AI represents a transformative opportunity to advance diagnostic accuracy, align computational approaches with human reasoning, and strengthen interdisciplinary collaboration. Realizing its clinical potential requires addressing substantial technical, ethical, and regulatory hurdles through collaborative innovation, rigorous evaluation, and proactive governance.

2. Landscape of Multimodal Generative AI Models in Clinical Diagnostics

A diverse set of multimodal generative AI models has recently emerged, each characterized by distinct architectures, strengths, and clinical capabilities. Med-PaLM M extends Google's PaLM language model through fine-tuning on medical datasets, employing transformer-based architectures with multimodal attention mechanisms to integrate textual data, medical imaging, and structured clinical information, thereby supporting tasks such as medical question answering and diagnostic reasoning [39]. LLaVA-Med adapts the Large Language and Vision Assistant (LLaVA) framework for clinical contexts by combining vision transformers specialized in medical image interpretation with language models, enabling effective joint analysis of radiology reports and their corresponding images [44]. BiomedGPT further broadens the scope of multimodal integration by incorporating genomic sequences, medical literature, protein structures, and clinical notes, using modality-specific encoders and cross-attention mechanisms to perform complex tasks such as biomedical entity recognition, hypothesis generation, and personalized treatment planning [45]. Finally, BioGPT-ViT combines Vision Transformer (ViT) capabilities for imaging analysis with GPT-based text processing, demonstrating utility in medical image captioning, visual question answering, and multimodal clinical decision support systems that align imaging data with electronic health records [27]. Collectively, these models represent complementary approaches toward unifying heterogeneous medical modalities, enabling both specialized and generalized diagnostic applications.

The conceptual design of a representative multimodal diagnostic AI architecture is illustrated in Figure 3. In this framework, modality-specific encoders—such as vision transformers for medical images, transformers for clinical text, 1D CNN or RNN modules for signals, and multilayer perceptrons for tabular data—are integrated via a cross-modal fusion layer (e.g., co-attention mechanisms). Task-specific heads, such as classifiers for diagnostic prediction or decoders for automated report generation, are appended to this shared latent space, while auxiliary modules for explainability (e.g., attribution maps, rules) and uncertainty calibration ensure clinical transparency and reliability.

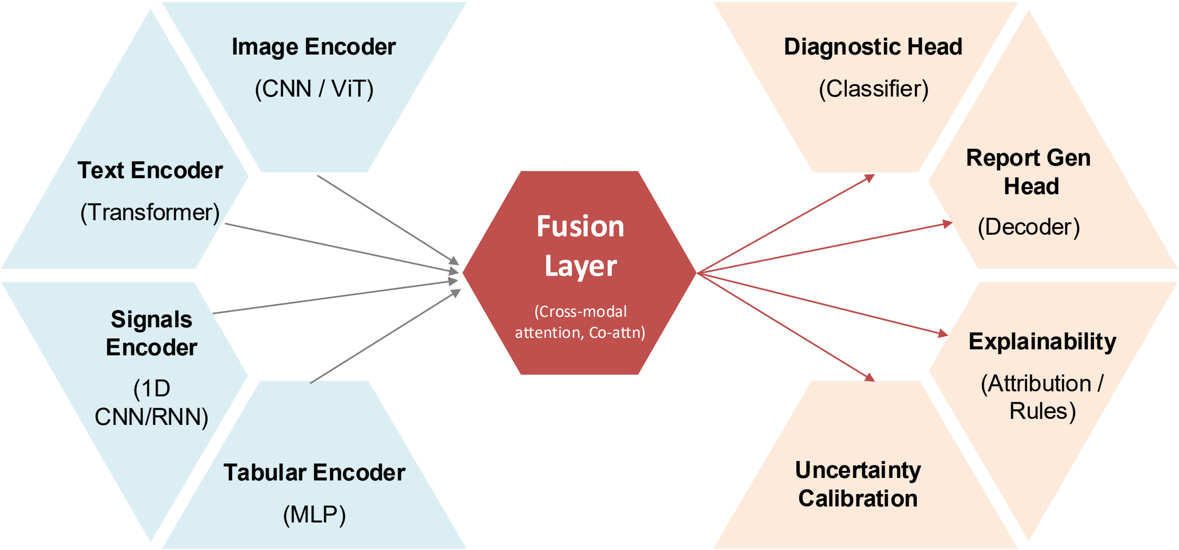


Figure 3. Reference multimodal architecture with modality-specific encoders (images, text, signals, tabular) feeding a fusion module and task-specific heads (classification, report generation), augmented by explainability and uncertainty calibration components; representative design patterns and clinical motivations are surveyed in recent multimodal reviews [46–49].

A structured comparison of the major multimodal generative AI models is presented in Table 2, which highlights their distinctive integrated modalities, clinical applications, strengths, and known limitations. Med-PaLM M demonstrates robust natural language interaction and domain knowledge but remains limited in terms of real-world validation [29]. LLaVA-Med shows high performance in CT and X-ray interpretation but underperforms in ultrasound analysis [41]. BiomedGPT excels at broad biomedical integration across genomics, proteins, literature, and clinical notes, though its interpretability is constrained by model complexity [39]. BioGPT-ViT effectively merges imaging and textual modalities for clinical reasoning support and analytics, yet faces risks of hallucinations in interpretation that may undermine clinical trust [40]. Together, these models underscore both the promise and the persistent barriers of multimodal AI in advancing diagnostic practice.

Table 2. Capabilities and limitations of representative multimodal generative AI models in clinical diagnostics.

Model Name	Integrated Data Modalities	Clinical Use-cases	Key Strengths	Known Limitations
Med-PaLM M	Medical images, clinical text, structured patient data	Medical QA, diagnostic support	Rich medical knowledge, natural language interaction	Limited real-world clinical validation [29]
LLaVA-Med	Medical images, radiology reports	Radiology reporting, visual QA	Effective CT, X-ray interpretation	Reduced accuracy in ultrasound [41]
BiomedGPT	Genomics, protein structures, literature, clinical notes	Drug discovery, literature synthesis, hypothesis generation	Multimodal biomedical data integration	Complexity in interpretability [39]
BioGPT-ViT	Medical imaging, clinical text, electronic health records	Image captioning, visual QA, clinical analytics	Integrated image-text analysis, clinical reasoning support	Potential hallucinations in interpretations [40]

Multimodal generative AI models differ substantially in their capacity to integrate diverse data modalities, and these differences directly shape their clinical effectiveness and scope of application. Med-PaLM M and LLaVA-Med specialize in combining medical imaging with clinical text, streamlining radiology workflows by providing preliminary interpretations and automated diagnostic recommendations that reduce clinician workload while improving diagnostic accuracy [39,44]. BiomedGPT expands the integration horizon by incorporating genomic sequences, temporal data, medical literature, protein structures, and clinical notes, thereby enabling comprehensive disease profiling particularly suited for complex conditions that require multidimensional biological, clinical, and imaging insights [45]. BioGPT-ViT, by merging Vision Transformer-based image analysis with GPT-driven language processing, has proven effective in tasks such as image captioning, visual question answering, and multimodal clinical decision support, excelling in real-time scenarios where clinical text and imaging data must be synthesized simultaneously to guide decision-making [27]. Collectively, these models demonstrate the capacity of multimodal AI to improve diagnostic workflows across radiology, oncology, and other specialties. For example, Med-PaLM M and LLaVA-Med have shown utility in automating radiology reports and facilitating early detection of abnormalities [39,44], while BiomedGPT enhances oncology workflows by integrating imaging, molecular, and clinical findings into nuanced, individualized diagnostic recommendations [45]. BioGPT-ViT extends this further by enabling clinically relevant multimodal analytics that can provide detailed image-captioning and real-time reasoning support in complex diagnostic settings [27]. The breadth of these data modalities and their clinical applications are illustrated in Figure 4, which maps imaging, textual, physiologic, structured, and omics data streams to downstream diagnostic tasks such as detection, segmentation, report generation, prognosis, and triage.

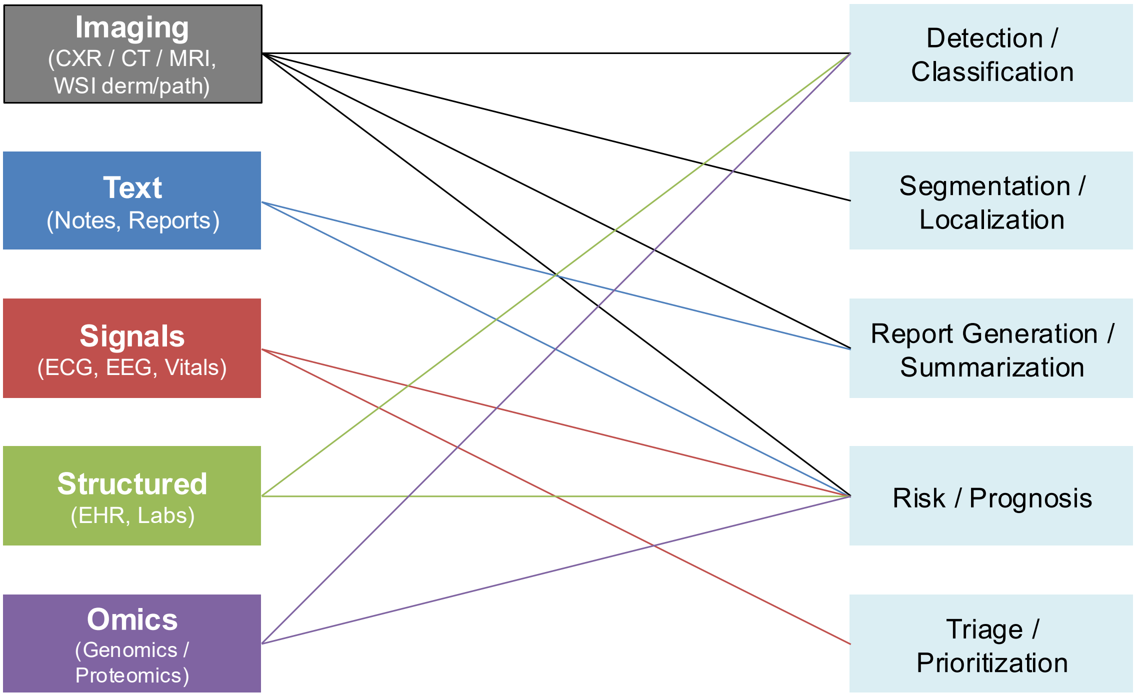


Figure 4. Common clinical data modalities (imaging, free-text/notes, physiologic signals, structured EHR, omics) mapped to key downstream tasks (detection/classification, segmentation/localization, report generation, prognosis/risk, triage), reflecting current multimodal biomedical AI practice [46,48].

Despite these promising advances, multimodal generative AI continues to face significant barriers that limit clinical adoption. The most fundamental challenge is data quality, as model performance depends on large, representative, and well-annotated multimodal datasets. Inadequate representation of minority populations introduces biases that can exacerbate existing disparities in healthcare delivery [29]. Equally critical are interpretability challenges: the opacity of decision-making in complex multimodal architectures undermines clinician confidence and trust, particularly in high-stakes diagnostic environments [50]. Resource intensity poses additional barriers, as training and deploying multimodal

doi:10.20944/preprints202508.1425.v2

models requires considerable computational infrastructure, which risks widening the gap between well-resourced health systems and underserved settings [24]. Furthermore, ethical and privacy issues are amplified when integrating multiple sensitive data types; the risks of patient re-identification and data misuse necessitate stringent safeguards and regulatory oversight [29]. Compounding these challenges, regulatory frameworks lag behind technological innovation, leaving clinical stakeholders uncertain about validation standards, approval pathways, and post-market monitoring requirements for multimodal AI systems [50].

Addressing these challenges requires methodological, infrastructural, and governance innovations. Priority directions include the development of enhanced interpretability methods capable of explaining multimodal reasoning processes in clinically meaningful ways; the application of federated learning to facilitate cross-institutional collaboration without compromising data privacy; and the creation of standardized data integration pipelines that harmonize heterogeneous modalities for robust clinical use. Large-scale prospective clinical trials are essential to validate efficacy across real-world patient populations, while interdisciplinary collaboration among clinicians, AI researchers, policymakers, and ethicists will be crucial to navigate the ethical and regulatory landscape. In sum, although multimodal generative AI has already demonstrated its potential to improve diagnostic accuracy, workflow efficiency, and personalized care across medical specialties, its long-term clinical translation depends on overcoming these technical, ethical, and regulatory hurdles through coordinated innovation and rigorous validation.

3. Multimodal LLM Design Approaches, Trade-offs, and Clinical Implications

The design of multimodal large language models (LLMs) for clinical diagnostics has rapidly advanced, with three dominant architectural paradigms—*tool use*, *grafting*, and *unification*—each offering distinct trade-offs between flexibility, integration depth, interpretability, and computational efficiency. Tool-use approaches equip an LLM to orchestrate specialized external models for non-text modalities such as images, signals, or genomics, thereby functioning as integrative coordinators that route queries to the most appropriate expert module before synthesizing the results. This modularity allows the incorporation of validated, domain-specific tools and supports rapid updates without retraining the entire system, though it introduces latency, dependency complexity, and only shallow cross-modal learning [39]. Grafting strategies connect pre-trained modality-specific encoders—such as vision transformers or clinical BERT variants—into an LLM backbone via adapters or fine-tuning, enabling joint intermediate processing of multimodal representations. These models achieve balanced performance across modalities with moderate resource demands, but their integration depth is limited and redundancy across modalities can reduce efficiency [27]. By contrast, unification strategies train a single end-to-end architecture on multiple modalities simultaneously, enabling deep cross-modal reasoning and holistic diagnostic insight by embedding imaging, textual, genomic, and structured data into shared latent spaces. Unified models streamline analysis and eliminate reliance on external tools, but they demand extensive multimodal datasets, vast computational resources, and careful balancing to avoid overfitting or performance collapse across modalities [50].

The comparative diagnostic value of these paradigms is illustrated in Figure 5, which highlights reported improvements in AUROC relative to unimodal baselines across diverse clinical tasks. These findings emphasize that while unified models tend to yield the largest performance gains in integrative reasoning tasks, grafting and tool-use strategies can outperform in contexts requiring modularity, rapid specialization, or reliance on expert subcomponents.

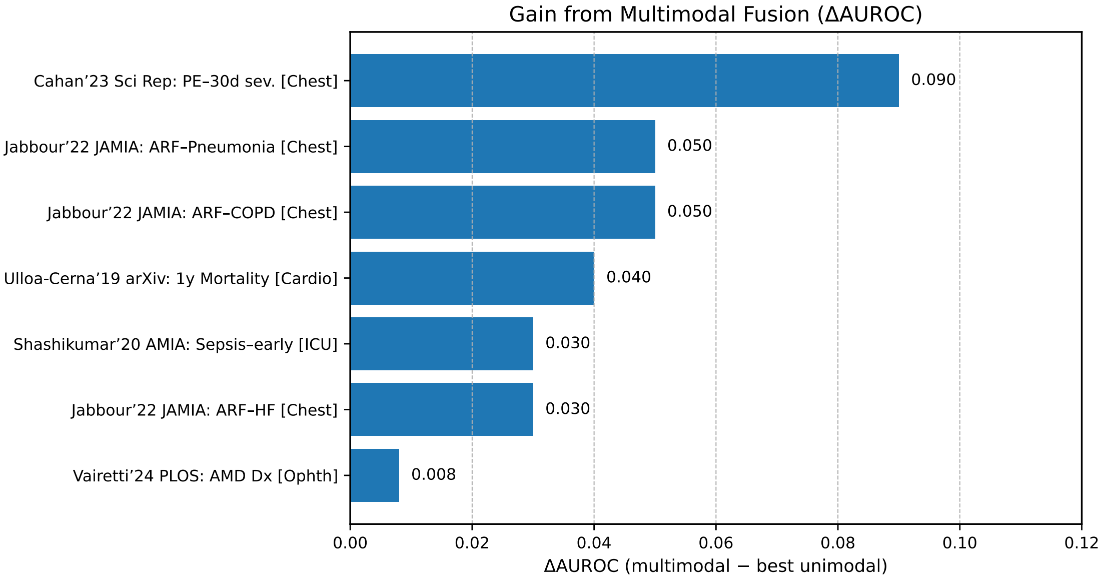


Figure 5. Reported improvements in AUROC (Δ vs. best unimodal baseline) across representative studies and clinical settings—acute respiratory failure diagnosis from chest radiographs+EHR [51], pulmonary embolism detection and severity risk using CT+EHR fusion [52,53], early sepsis detection from multimodal clinical data [54], and multimodal ophthalmic models for age-related macular degeneration [55].

3.0.1. Tool Use Approach

The tool-use methodology equips large language models (LLMs) with the ability to orchestrate external specialized tools or domain-specific models, thereby extending their capacity to process modalities beyond their native textual input. In this paradigm, the LLM serves as an integrative coordinator that determines when additional modality-specific expertise is required and invokes external systems accordingly, whether for image analysis, speech recognition, or genetic interpretation. The outputs from these specialized modules are subsequently integrated into the LLM’s reasoning process to provide comprehensive diagnostic insights or recommendations [56]. For instance, in radiological workflows, an LLM may request a convolutional neural network (CNN)-based model for lung nodule characterization and then combine these imaging results with textual patient records to generate a cohesive diagnostic report. This strategy offers considerable flexibility and modularity, as new or improved specialist modules can be incorporated without retraining the core model, while also capitalizing on extensively validated domain-specific tools to enhance diagnostic accuracy in specialized tasks. Nonetheless, tool-use approaches are not without drawbacks, as sequential invocation of multiple tools can introduce latency, the management of interdependencies among heterogeneous components increases system complexity, and the cross-modal learning achieved is often limited because integration occurs at a higher and more superficial representational level rather than through deeply unified embeddings.

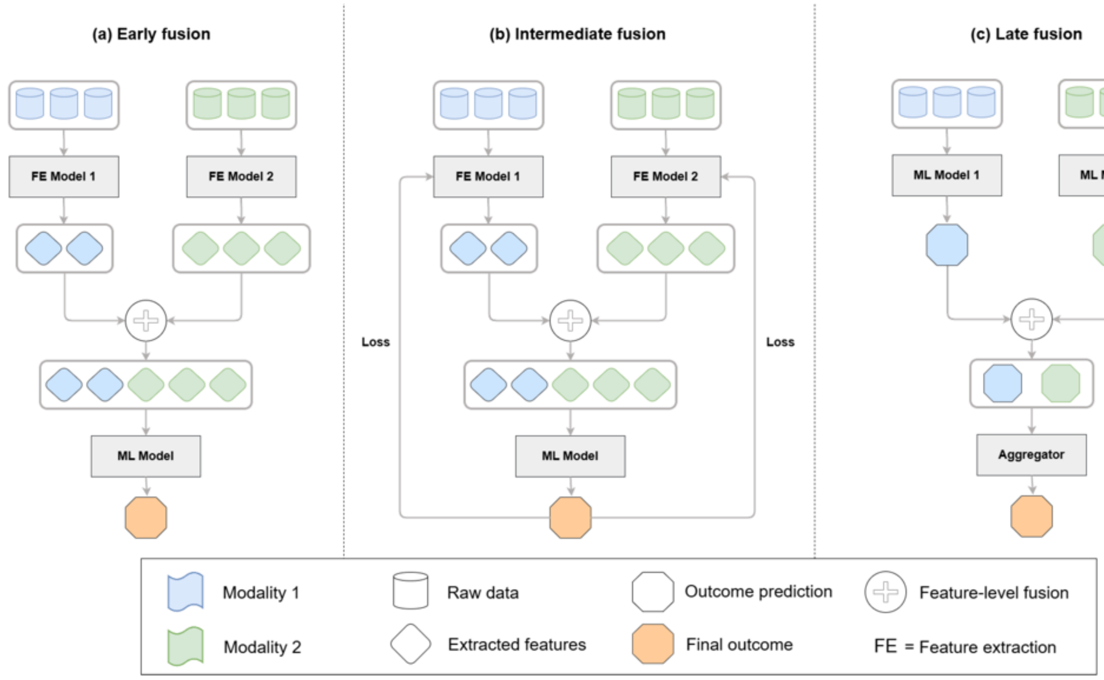
3.0.2. Grafting Approach

The grafting approach integrates pre-trained, modality-specific components directly into a foundational LLM via adapter layers or targeted fine-tuning, thereby allowing multimodal data to be jointly processed within the model’s representational hierarchy. By leveraging specialized encoders—such as vision transformers trained on imaging tasks or BERT-derived models adapted for clinical text—this strategy enables simultaneous interpretation of multiple modalities, such as pairing medical images with clinical notes to improve performance in pathology and radiology tasks [27]. Grafted models benefit from reusing mature pretrained networks, thereby reducing both the volume of training data required and the associated computational costs compared to building fully unified models from scratch. Moreover, they tend to achieve balanced performance across modalities, making them well suited for clinical scenarios in which moderate integration between modalities is sufficient to yield actionable insights. However, the approach remains constrained by limited depth of cross-modal interaction, which may prevent capture of subtle interdependencies between modalities, as well as by redundancy in representational learning, which can reduce efficiency and hinder scalability when expanding to additional data types.

These architectural approaches can be systematically categorized by their fusion strategies, which determine how different modalities are combined within a model. As illustrated in Figure 6, fusion can occur at the feature level (early fusion), within latent representations (intermediate fusion), or at the decision level (late fusion). Hybrid strategies that combine feature- and decision-level integration are also increasingly applied in biomedical AI, balancing the advantages of deep integration with practical considerations of efficiency and interpretability [47,48].

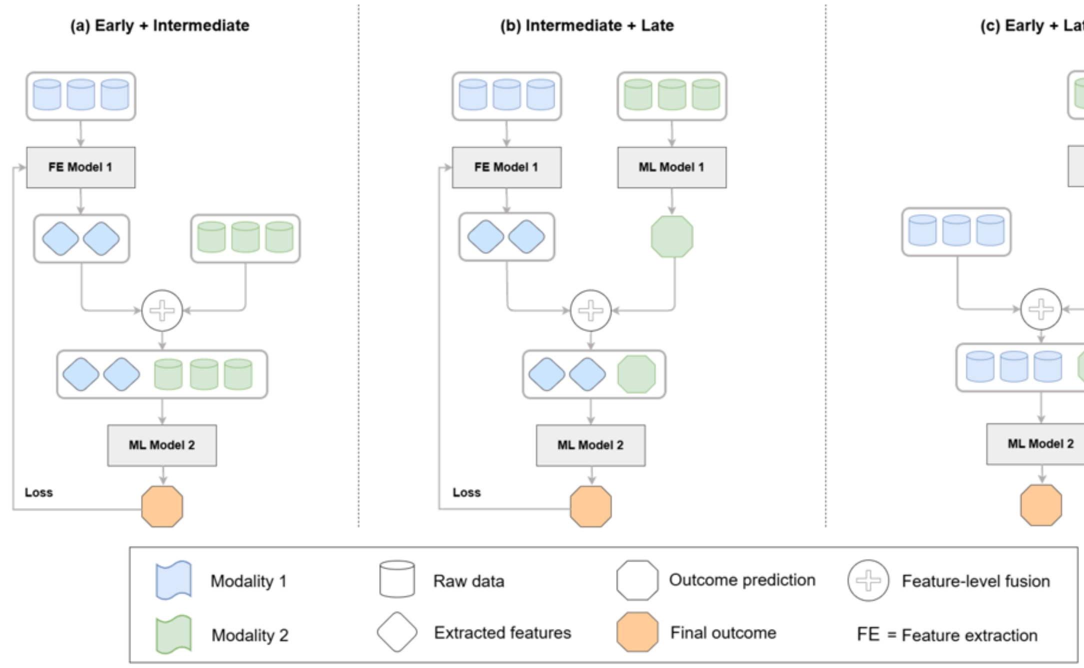
3.0.3. Unification Approach

Unified models represent the most ambitious paradigm, in which a single holistic model is trained to simultaneously ingest and integrate diverse modalities—including medical imaging, textual reports, genomics, and structured clinical data—within a shared representational framework. These systems rely on advanced architectural designs and complex attention mechanisms to achieve deep cross-modal learning, enabling robust internal representations that capture intricate relationships across data types. For example, a unified multimodal transformer may concurrently analyze radiological images, genomic profiles, and clinical notes to produce a diagnostic assessment enriched by highly integrated insights [29]. This design excels in capturing complex interdependencies that simpler approaches may miss, while also streamlining the analytical pipeline by reducing reliance on external modules, thereby improving overall coherence and reliability of outputs. However, unified models are computationally intensive, demanding substantial GPU resources and access to expansive, well-curated multimodal datasets. Their training is challenging, as performance must be carefully balanced across heterogeneous modalities, and risks such as overfitting or reduced generalization to real-world clinical variability remain significant obstacles. Despite these challenges, unified models hold the greatest promise for uncovering novel clinical insights and supporting deeply integrated diagnostic reasoning in multimodal healthcare applications.



(**a**) Multimodal fusion strategies.

Figure 6. *Cont.*



(**b**) Hybrid (mixed) fusion patterns.

Figure 6. Taxonomy of multimodal fusion strategies—early/feature-level, intermediate/latent, late/decision-level—and hybrid patterns; biomedical fusion reviews and case studies detail these patterns and their performance trade-offs in practice [47,48,52].

3.1. Comparative Evaluation of Multimodal LLM Strategies

The comparative evaluation of multimodal large language model (LLM) design strategies highlights distinct strengths and limitations that influence their suitability for different clinical contexts (Table 3). Unified models generally demonstrate superior performance in tasks requiring deeply integrated reasoning across heterogeneous modalities, leveraging their ability to capture complex interdependencies between imaging, textual, and structured clinical data. However, tool-use and grafting approaches can outperform unified systems in scenarios where specialized domain expertise, modular flexibility, or rapid iterative updates are required. Tool-use strategies excel by orchestrating validated external expert modules—such as convolutional networks for radiology—providing flexibility and high domain accuracy, albeit with challenges such as latency, dependency management, and only superficial integration [38,56]. Grafting approaches, by embedding pre-trained encoders for different modalities into an LLM backbone, deliver balanced performance and efficient use of computational resources, making them attractive for clinical settings where moderate multimodal integration is sufficient. Yet, their limited depth of cross-modal reasoning and redundancy across modality-specific representations restrict their scalability. In contrast, unified models streamline the diagnostic pipeline by training a single architecture across modalities, offering deep cross-modal reasoning and holistic insights, but at the expense of computational cost, complex training requirements, and higher risk of overfitting [50]. Interpretability further distinguishes these strategies: tool-use approaches provide clearer transparency regarding which module contributes to a decision, while unified models, though powerful, often deliver more opaque reasoning processes that can hinder clinician trust and regulatory acceptance.

Table 3. Comparative Summary of Multimodal LLM Design Strategies

Strategy	Methodology	Representative Models	Clinical Examples	Key Strengths	Limitations
Tool Use	External tool orchestration by LLM	GPT-4 with plugin architecture [56]	Radiology image analysis, lung nodule evaluation	Flexibility, modularity, high domain accuracy	Latency, complexity in management, shallow integration
Grafting	Pre-trained encoders connected to LLM	VisualBERT, CLIP [27]	Pathology slide analysis, histology-text integration	Efficient resource use, balanced performance	Limited deep integration, redundancy of features
Unification	Single model integrated training	Multimodal Transformers [29]	Comprehensive diagnostics (images, EHR, genomics)	Deep cross-modal reasoning, streamlined pipeline	High computational cost, complex training, risk of overfitting

The relative diagnostic value of these paradigms is further illustrated in Figure 8, which presents multimodal AUROC performance across representative clinical tasks compared against the best unimodal comparators. These results underscore that unified models tend to yield the largest gains in contexts requiring deep cross-modal reasoning, while tool-use and grafting strategies retain significant value in specialized or resource-constrained settings where modularity, interpretability, or efficiency take precedence.

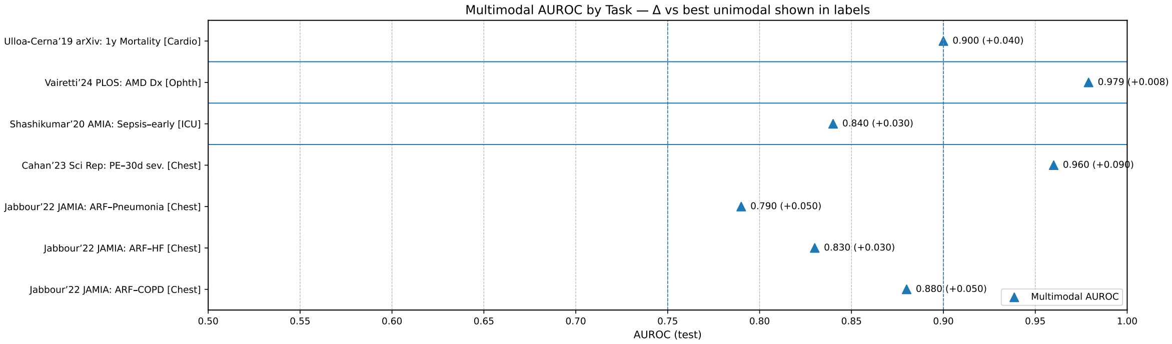


Figure 7. Multimodal AUROC by clinical task with per-bar callouts indicating the margin over the best unimodal comparator; exemplars include chest radiograph+EHR for acute respiratory failure [51], CT+EHR for pulmonary embolism detection and 30-day risk [52,53], multimodal sepsis detection [54], and multimodal ophthalmic diagnosis [55].

3.2. Specialization and Generalization Trade-offs

An important dimension in the design of multimodal generative AI systems is the balance between specialization and generalization, as each orientation carries distinct advantages and trade-offs for clinical practice. Specialized models are typically optimized for narrowly defined diagnostic tasks, achieving high accuracy and strong domain-specific interpretability. Such models excel in targeted applications like oncology, where precision and transparency are critical for tasks such as cancer diagnosis or prognosis prediction [33,38]. In contrast, generalized models offer broad applicability across diverse clinical scenarios, from primary care to emergency medicine, enabling versatility in settings that require rapid, adaptable diagnostic support. However, this breadth comes with potential reductions in peak performance compared to specialized counterparts [29]. These trade-offs extend beyond performance to encompass interpretability, computational requirements, and deployment complexity. Specialized systems are often easier to interpret, require fewer computational resources, and can be more straightforward to deploy within narrowly defined workflows. By contrast, generalized models demand greater computational infrastructure, pose challenges in explaining complex multimodal reasoning across varied domains, and require more intricate integration into heterogeneous clinical workflows [45,57]. The comparative characteristics of these two design orientations are summarized in Table 4, which outlines their respective strengths, limitations, and implications for clinical implementation.

Table 4. Comparison of Specialized vs. Generalized Multimodal AI Models in Diagnostics

Aspect	**Specialized Models**	**Generalized Models**	**Performance Implications**
Clinical Focus	Oncology, specific diseases, targeted pathology	Primary care, emergency medicine, broad diagnostics	High accuracy (specialized); broad versatility (generalized)
Examples	PathChat for breast cancer pathology [33]	PathChat in diverse pathology domains [33]	Specialized: precise domain accuracy; Generalized: moderate across multiple domains
Interpretability	Clear domain-specific explanations	Complex reasoning across multiple tasks	Specialized: easier regulatory approval; Generalized: greater complexity in validation
Resource Efficiency	Lower resources for specific tasks	Higher computational requirements for diverse capabilities	Specialized: efficient, task-specific; Generalized: resource-intensive for training and deployment
Deployment Complexity	Simpler for focused clinical scenarios	Complex integration in varied workflows	Specialized: straightforward integration; Generalized: complex implementation

3.3. Future Directions and Research Priorities

Moving forward, advancing multimodal generative AI in diagnostics requires systematically addressing ongoing technical, ethical, and regulatory challenges. Improving interpretability remains a top priority, as clinicians must be able to trust and understand the reasoning behind AI-assisted decisions in high-stakes clinical contexts. Strategies such as federated learning and transfer learning will be essential for managing data scarcity, enabling institutions to collaborate without compromising patient privacy while broadening the representativeness of training datasets. Equally important is the development of robust validation frameworks, including large-scale prospective clinical trials, to demonstrate reproducibility and generalizability across diverse patient populations. Hybrid approaches that integrate aspects of tool-use, grafting, and unification may offer a pragmatic balance, combining modularity and efficiency with deeper integration, thereby maximizing clinical utility. Emphasis on causal learning, explainable AI methods, and standardized benchmarking datasets will be vital to achieving regulatory acceptance and facilitating widespread clinical adoption [29,58]. In conclusion, careful consideration of design strategies, performance trade-offs, and practical implementation challenges is necessary for translating multimodal generative AI into meaningful clinical applications, ensuring improvements not only in diagnostic accuracy but also in workflow efficiency and patient outcomes.

4. Clinical Applications of Multimodal Generative AI: Radiology, Pathology, Dermatology, and Ophthalmology

Multimodal generative artificial intelligence (AI) has shown significant potential to revolutionize medical diagnostics across various specialties by integrating diverse data sources and enhancing clinical decision-making. This section reviews prominent clinical applications, capabilities, comparative benefits, and existing challenges of multimodal AI within radiology, pathology, dermatology, and ophthalmology.

4.1. Radiology Applications

Radiology has embraced multimodal AI extensively, integrating imaging data (CT, MRI, PET) with electronic health records (EHR) and clinical notes to achieve improved diagnostic accuracy and streamlined workflows. Multimodal deep learning approaches have demonstrated superior diagnostic capabilities compared to traditional, unimodal methods, notably in the diagnosis of neurodegenerative diseases. For instance, a multimodal classifier combining FDG-PET and MRI achieved enhanced diagnostic accuracy in Alzheimer's disease classification, leveraging complementary imaging modalities to improve sensitivity and specificity [31]. Moreover, the integration of vision-language models facilitates tasks such as visual question answering (VQA) and automated report generation from imaging data, significantly improving the efficiency and interpretability of radiological workflows [25]. The ROCO (Radiology Objects in COntext) dataset exemplifies the utility of multimodal benchmarks, offering comprehensive data for developing and validating models capable of tasks such as radiographic image captioning and text-conditioned image retrieval. Despite these advancements, challenges remain in data quality, model interpretability, and integration within existing clinical systems. Large-scale validation and regulatory approval processes also represent significant hurdles for practical implementation [44,59].

4.2. Pathology Applications

In pathology, multimodal AI systems that integrate histopathology images, genomic data, and textual clinical information have significantly advanced precision diagnostics and personalized treatment planning. The PathChat model is a notable example, combining a specialized pathological image encoder with a language model, achieving high diagnostic accuracy (up to 89.5%) on expert-curated multimodal tasks, significantly surpassing single-modality systems [33]. Similarly, the CONCH model demonstrates potential for rare disease diagnostics, utilizing large-scale multimodal datasets for zero-shot pathology classification [33].

PathVQA and related visual question answering datasets support the training and evaluation of AI systems in pathology, testing their ability to integrate visual and textual diagnostic reasoning. Yet, interpretability and usability remain significant challenges; black-box nature and limited transparency of complex models hinder clinician trust and acceptance. Efforts such as explainable AI (XAI) methods (e.g., attention maps, SHAP values) and federated learning approaches for privacy-preserving collaboration across institutions aim to address these barriers [60,61].

4.3. Dermatology Applications

Dermatology benefits from multimodal generative AI through enhanced diagnostic accuracy in skin lesion classification and early detection of skin cancers by integrating dermatoscopic images, clinical images, and patient metadata. Multimodal deep learning models demonstrate superior performance to unimodal approaches, particularly in distinguishing malignant from benign lesions, reducing unnecessary biopsies, and improving overall diagnostic confidence [34,37]. Conversational AI capabilities further improve patient-provider communication, aiding patients in understanding complex medical conditions and treatment options through accessible, natural-language explanations of diagnostic findings and management strategies [35].

Generative AI techniques address data scarcity by creating realistic synthetic dermatological images, supporting model training and educational data generation for rare conditions. Despite these strengths, data bias, model interpretability, and integration into clinical workflows present challenges, demanding continued development of standardized data protocols, rigorous validation, and advanced interpretability methods [35].

4.4. Ophthalmology Applications

Ophthalmology has rapidly adopted multimodal AI to improve diagnostic accuracy and disease monitoring, particularly in diabetic retinopathy, glaucoma, and age-related macular degeneration. Multimodal generative AI models integrating fundus photographs, OCT scans, and clinical records demonstrate enhanced diagnostic precision and disease progression prediction compared to single-modality analyses, achieving notable area-under-curve (AUC) scores exceeding 0.80 [36,37]. These systems facilitate earlier and more accurate diagnosis, potentially improving patient outcomes significantly.

Synthetic image generation capabilities further augment training datasets and aid medical education. However, practical implementation challenges include data standardization, model interpretability, clinical workflow integration, and ethical considerations regarding synthetic data use and privacy compliance [40,42].

4.5. Comparative Analysis Across Specialties

A comparative evaluation (Table **??**) highlights varying degrees of adoption, effectiveness, and challenges across radiology, pathology, dermatology, and ophthalmology. Radiology and pathology have experienced relatively greater adoption due to established digital data frameworks and imaging standardization. Dermatology and ophthalmology are swiftly advancing, leveraging recent digitization and generative AI techniques. Shared challenges across specialties include data integration complexity, interpretability concerns, data quality and bias risks, and regulatory hurdles. Standardized data protocols, enhanced model interpretability methods, and robust validation frameworks are critical for overcoming these barriers.

4.6. Future Directions and Clinical Integration

Future multimodal generative AI research should prioritize addressing these shared challenges through advanced interpretability techniques, federated learning frameworks for privacy preservation, standardized data protocols, and rigorous prospective clinical validation. Emphasis on interdisciplinary collaboration between AI researchers, clinicians, and regulators will be crucial for successful clinical translation. Developing robust governance and evaluation standards, alongside continuous monitoring and recalibration of AI models, will ensure sustained accuracy and clinical relevance across diverse patient populations and healthcare settings.

In summary, multimodal generative AI offers profound transformative potential across radiology, pathology, dermatology, and ophthalmology, promising significant improvements in diagnostic precision, personalized care, and clinical workflow efficiency. However, careful management of data quality, ethical considerations, model interpretability, and clinical integration barriers is essential to fully realize the clinical impact of these powerful diagnostic technologies.

5. Benchmark Datasets and Evaluation in Multimodal Generative AI Diagnostics

5.1. Overview and Importance

Multimodal generative artificial intelligence (AI) has experienced substantial growth in diagnostic applications across medical specialties, driven by comprehensive benchmark datasets that integrate imaging, textual, genomic, and clinical data. These datasets underpin advancements in model performance, facilitate methodological innovations, and enable rigorous validation of AI systems, significantly enhancing their diagnostic utility and generalizability. However, dataset-related challenges, including annotation consistency, diversity, quality, and ethical considerations, continue to impact AI development and clinical translation [29,43].

5.2. Benchmark Datasets for Radiology

Radiology has notably benefited from multimodal AI benchmarks such as the MIMIC-CXR dataset and ImageCLEF challenges. The MIMIC-CXR dataset comprises over 377,000 chest X-rays paired with radiological reports, facilitating tasks such as abnormality detection, disease classification, and automated report generation [41]. ImageCLEF further extends radiology multimodal research by hosting various tasks, including visual question answering, captioning, and image-text retrieval, providing a robust evaluation platform for multimodal models [44]. Representative AI systems, including VisualBERT and transformer-based architectures, leverage these datasets to demonstrate state-of-the-art performance, such as achieving AUC scores exceeding 0.95 for disease classification and high ROUGE scores in report generation tasks [44,62]. However, limitations such as single-institution bias in MIMIC-CXR, variability in annotation quality, and limited generalization across diverse patient populations highlight the need for larger, more representative datasets and standardized evaluation frameworks (Table 5). These challenges necessitate rigorous clinical validation to ensure AI model performance translates effectively into clinical practice [41,44].

Table 5. Key multimodal benchmarks & datasets for diagnostics (tasks, modalities, scale, metrics)

Aspect	MIMIC-CXR	ImageCLEF
Tasks Supported	Disease classification, abnormality detection, automated report generation [41]	Visual question answering, captioning, multi-label classification, image retrieval [44]
Modalities	Chest X-rays, radiology reports	Multi-modality images (X-rays, CT), text annotations
AI Models	VisualBERT, transformer models, graph-enhanced models [44]	Transformer-based, multi-task learning approaches [62]
Performance	High accuracy (AUC >0.95), strong report generation metrics (ROUGE-L >0.74)	Task-specific varied performance across multiple benchmarks
Limitations	Single-institution bias, annotation variability [41]	Smaller scale, annotation inconsistencies, clinical representativeness issues

5.3. Benchmark Datasets for Pathology

Pathology benefits from specialized multimodal datasets such as PathVQA, designed explicitly for histopathology image question-answering (QA). PathVQA pairs whole-slide images with diagnostic questions, facilitating fine-grained interpretative capabilities in AI systems [60]. Models like PathChat have leveraged these datasets to achieve high diagnostic accuracy, notably attaining accuracies of up to 89.5% when multimodal data are integrated, significantly outperforming unimodal methods [33]. The development of synthetic datasets, including those generated through Generative Adversarial Networks (GANs), further addresses data scarcity issues, enabling robust training of diagnostic AI models for rare pathologies [63,64]. Nevertheless, synthetic datasets face challenges regarding clinical validity, realism, and ethical implications. The realism of generated question-answer pairs and synthetic images must be carefully validated by clinical experts to ensure their suitability for diagnostic applications and training purposes. Moreover, maintaining ethical standards, ensuring patient privacy, and avoiding potential biases inherent in the original training data remain critical considerations [32,43].

5.4. Benchmark Datasets for Dermatology and Ophthalmology

Dermatology and ophthalmology have rapidly adopted multimodal benchmarks, combining images (dermatoscopic, fundus photography, OCT) with clinical metadata. Multimodal models have significantly improved lesion classification accuracy in dermatology, achieving high sensitivity and specificity compared to single-modality approaches [34,35]. Ophthalmology has similarly leveraged multimodal datasets to enhance diagnostic capabilities in diabetic retinopathy and glaucoma monitoring, achieving area-under-curve (AUC) scores exceeding 0.80 through the integration of fundus images, OCT scans, and clinical records [36,37]. Challenges in these fields parallel those observed in radiology and pathology, including data standardization, interpretability of AI models, ethical use of patient-derived data, and generalizability across diverse patient populations. Future directions emphasize larger, more diverse multimodal datasets, improved model interpretability, and robust clinical validation to ensure safe and effective clinical implementation.

5.5. Synthetic Image and Text Generation for Rare Conditions

The use of synthetic data, generated through GANs, diffusion models, and large language models (LLMs), addresses critical data scarcity issues in rare conditions and facilitates privacy-preserving collaboration. GANs and diffusion models successfully generate realistic synthetic medical images, augmenting rare disease datasets and improving diagnostic AI performance. Synthetic images have been validated by expert clinicians, achieving realism sufficient for training and clinical use [27,65]. LLM-generated synthetic clinical narratives, paired with medical images, further support medical education, training, and research into rare disease diagnostics, offering scenarios that might be rare or ethically challenging to capture in real patient data [29,32]. However, clinical validity, data quality, ethical concerns, and potential biases introduced by synthetic data require rigorous evaluation and standardization. Ensuring synthetic data accurately reflects clinical realities, maintaining patient privacy, and establishing clear regulatory frameworks are critical for responsible implementation. Ongoing research into advanced evaluation metrics, federated learning, and ethical guidelines is crucial for addressing these challenges and maximizing the utility of synthetic multimodal data in clinical diagnostics [43,56].

5.6. Future Trends and Research Priorities

Emerging trends in multimodal benchmark datasets for medical AI emphasize creating larger, diverse, and representative datasets with standardized data collection and annotation protocols. Synthetic data generation techniques, privacy-preserving methods like federated learning, and benchmark datasets for rigorous model evaluation are rapidly advancing. Standardization of data formats and clinical integration approaches will facilitate practical deployment. Increased interdisciplinary collaboration, improved ethical frameworks, rigorous validation protocols, and clear regulatory guidelines are critical next steps for realizing the full clinical potential of multimodal generative AI technologies across medical specialties [29,43,44].

In conclusion, benchmark datasets and synthetic data approaches significantly advance multimodal AI in diagnostics, offering transformative clinical capabilities across various medical specialties. Addressing existing data quality, interpretability, ethical, and regulatory challenges remains essential for successful translation into clinical practice and sustained improvement of patient outcomes.

6. Synthetic Multimodal Data Generation and International Collaboration in Rare Disease Research

6.1. Synthetic Multimodal Data Generation Methods

Synthetic multimodal data has become an indispensable tool for addressing data scarcity in rare disease research, offering ways to generate clinically meaningful and privacy-preserving datasets that extend beyond the limitations of real-world patient data. Several methodologies have been developed, each with unique benefits and limitations, and together they form the foundation for building robust synthetic datasets that can support diagnostic model training, educational applications, and international collaboration.

6.1.1. Generative Adversarial Networks (GANs)

Generative adversarial networks (GANs) have been particularly successful in producing realistic synthetic medical images, clinical narratives, and genomic data. Built upon a dual-network architecture where a generator produces synthetic samples and a discriminator evaluates their realism, GANs iteratively refine outputs until they are indistinguishable from authentic data. This paradigm has demonstrated value for augmenting rare disease datasets, strengthening model robustness, and providing synthetic material in situations where real patient data is sparse or unavailable [32,63]. By creating diverse, high-quality outputs without directly exposing sensitive information, GANs enhance both the breadth and depth of training datasets, although concerns remain around the clinical validity of generated data and the significant computational demands required to maintain fidelity.

6.1.2. Variational Autoencoders (VAEs)

Variational autoencoders (VAEs) offer an alternative generative framework by learning compressed latent representations of input data and decoding them into new synthetic samples. VAEs have been successfully applied to produce synthetic electronic health records and medical imaging, demonstrating their ability to capture underlying data distributions while preserving essential clinical properties [32]. The advantages of VAEs include efficiency in data representation and privacy preservation, as well as reduced burden on original datasets. However, their outputs can sometimes oversimplify complex clinical patterns, raising concerns about representation fidelity and specificity when applied to rare and heterogeneous disease populations.

6.1.3. Diffusion Models

Diffusion models have emerged as a powerful class of generative methods, producing synthetic medical images through iterative processes of noise addition and removal. They have shown particular utility in generating realistic neuroimaging data for rare neurological diseases where authentic datasets are extremely limited. While diffusion models produce highly realistic outputs and support improved diagnostic model training, they are computationally intensive and demand substantial expertise for successful implementation, with training complexity representing a barrier for widespread adoption [29,65].

Table 6. Synthetic data generation methods for rare-disease research: modalities, use-cases, advantages, limitations

Method	Data Modalities / Types	Clinical Utility / Use-cases	Key Advantages	Challenges / Limitations
Generative Adversarial Networks (GANs) [66]	Imaging (histopathology, radiology), genomics, narratives	Dataset augmentation, diagnosis support, rare condition modeling, educational uses	Realistic outputs, enhanced data diversity	Clinical validity concerns, computational demand
Diffusion Models [67,68]	Imaging (MRI, PET)	Rare neurological disease modeling, high-quality image synthesis	Realistic, high-quality medical imaging	High computational requirements, complex training
Variational Autoencoders (VAEs) [69]	Imaging, clinical records	Dataset expansion, privacy preservation, efficient data representation	Data efficiency, privacy protection, reduced data burden	Representation fidelity, dataset specificity, potential oversimplification
Large Language Models (LLMs) [70,71]	Clinical narratives, paired images	Medical training, hypothesis generation, synthetic patient scenarios, educational data	Rich narrative detail, flexible clinical scenario creation	Hallucination risk, potential inaccuracies, realism validation
Differential Privacy & Federated Learning [72,73]	Multimodal medical data	Cross-institutional and international collaboration, privacy assurance	Strong privacy guarantees, regulatory alignment, supports data sharing without transfer	Implementation complexity, model synchronization challenges

6.1.4. Large Language Models (LLMs) for Clinical Narratives

Large language models such as GPT-4 extend synthetic data generation into the textual domain by producing clinical narratives that closely resemble authentic patient records. When paired with synthetic images, these narratives allow for the creation of multimodal synthetic scenarios that serve educational purposes, hypothesis generation, and preliminary experimentation in rare disease contexts. This significantly enhances the availability of diverse datasets, particularly in underrepresented conditions, while mitigating privacy risks [29,32]. The challenge lies in ensuring narrative realism and mitigating risks of hallucination or factual inaccuracies.

6.1.5. Differential Privacy and Federated Learning

Complementary approaches such as differential privacy and federated learning focus not on generating data but on preserving privacy and enabling collaborative research. Differential privacy introduces calibrated noise to data or model parameters to mathematically guarantee confidentiality, while federated learning allows distributed model training across institutions without direct patient data exchange. Together, these methods facilitate international collaboration by addressing regulatory and ethical challenges associated with cross-border data sharing, ensuring that synthetic or aggregated multimodal datasets remain both useful and compliant [40,56].

Table 7. Overview of Synthetic Multimodal Data Generation Methods for Rare Disease Research

Method	Data Types Generated	Clinical Applications	Key Advantages	Limitations
Generative Adversarial Networks (GANs) [66]	Imaging, genomic, clinical narratives	Rare disease diagnosis, training augmentation	Realistic outputs, data augmentation capability	Clinical validity concerns, computational intensity
Variational Autoencoders (VAEs) [69]	Imaging, clinical records	Dataset expansion, privacy protection	Efficient data representation, preserves privacy	Representation fidelity, dataset specificity
Diffusion Models [67,68]	Imaging (MRI, PET)	Rare neurological disease imaging	High-quality image synthesis, data augmentation	Computational demands, complex training
Large Language Models (LLMs) [70,71]	Clinical narratives, paired images	Educational scenarios, hypothesis generation	Rich narrative detail, flexible scenario creation	Potential hallucinations, realism validation
Differential Privacy and Federated Learning [72,73]	Multimodal medical data	Cross-border collaboration, privacy preservation	Strong privacy guarantees, regulatory alignment	Complexity, model synchronization challenges

6.2. Enabling International Collaboration in Rare Disease Research

Beyond data generation, synthetic multimodal methods play a pivotal role in enabling international collaboration by mitigating regulatory barriers, enhancing data availability, and providing standardized outputs for shared research. Synthetic datasets can be shared across borders without exposing real patient information, thereby addressing the stringent privacy constraints imposed by frameworks such as GDPR and HIPAA. This reduces ethical risks while enabling more rapid and secure collaboration among institutions worldwide [29,56]. In addition, synthetic datasets augment rare disease cohorts by increasing both the volume and diversity of available cases, while standardized outputs improve interoperability and ensure that multimodal data generated across institutions can be seamlessly integrated into joint studies [43]. Importantly, the availability of synthetic datasets also facilitates pre-research exploration, allowing collaborators to test analytical workflows, train baseline models, and plan study designs before formal data-sharing agreements are finalized, thereby accelerating the pace of research and improving patient outcomes [56].

6.3. Practical Implementations and Case Studies

Several international initiatives already demonstrate the utility of synthetic multimodal data in both rare disease research and broader clinical contexts. The MIMIC series, while focused on critical care rather than rare diseases, has pioneered de-identified and synthetic dataset sharing, establishing benchmarks and facilitating collaborative research globally [44]. Alzheimer's research has successfully employed synthetic multimodal datasets, including neuroimaging and clinical narratives, to enhance diagnostic accuracy, monitor disease progression, and improve modeling of neurodegenerative trajectories, directly informing applications in rare neurological conditions where authentic datasets remain scarce [36,65]. Similarly, synthetic genomic and phenotypic datasets have enabled international collaboration in rare genetic disorder research, providing privacy-preserving synthetic cohorts and pedigrees that support diagnostic model development and therapeutic investigations [43,74]. These examples, summarized in Table 8, underscore the clinical and research value of synthetic data across diverse contexts while also highlighting persistent challenges around realism validation, generalizability, and dataset bias.

Table 8. Case studies of synthetic data applications in rare disease and critical care research

Case Study	Synthetic Data Types	Clinical/Research Impact	Limitations / Challenges
MIMIC Series [44]	Multimodal critical care data (images, clinical records)	Enabled international collaboration, established benchmarks, facilitated collaborative research	Not specific to rare diseases, limited representativeness
Alzheimer's Disease [36,65]	Synthetic MRI, PET, clinical narratives	Improved early detection, diagnostic accuracy, and disease progression modeling	Realism validation, clinical validity concerns, applicability to diverse populations
Rare Genetic Disorders [43,74]	Synthetic genomic data, phenotypic profiles	Enhanced diagnostic model development, enabled privacy-preserving collaboration	Validation complexity, potential dataset biases

6.4. Future Research Priorities

Future research in synthetic multimodal data generation should focus on enhancing realism and clinical fidelity, improving interpretability and transparency of generative methods, and addressing persistent ethical and regulatory concerns. The integration of advanced architectures such as diffusion models and hybrid generative frameworks with federated learning protocols can ensure both high-quality outputs and strong privacy protection. Rigorous expert validation protocols are critical for assessing realism, mitigating bias, and ensuring synthetic datasets are safe for clinical and research applications. Furthermore, sustained interdisciplinary collaboration—spanning AI researchers, clinicians, ethicists, and regulatory authorities—will be essential for standardizing methodologies, establishing governance frameworks, and fostering responsible adoption. Through these efforts, synthetic multimodal data can maximize its potential to advance rare disease diagnostics, support international research, and deliver equitable healthcare innovations globally.

7. Critical Assessment of Synthetic Data Approaches for Rare Disease Research

7.1. Data Quality and Clinical Validity

Ensuring the clinical validity and quality of synthetic multimodal medical data remains paramount, particularly in rare disease research where diagnostic precision is critical. Validation studies comparing AI models trained on synthetic versus real patient data are crucial for confirming that synthetic datasets accurately reflect the complexity and subtleties inherent in genuine clinical scenarios [43]. This task is notably challenging in rare diseases, where limited original data and incomplete disease understanding can complicate the establishment of robust validation frameworks. The accuracy, clinical relevance, and representativeness of synthetic datasets must therefore be rigorously assessed by domain experts to avoid misleading conclusions in diagnostics and treatment planning.

7.2. Realism, Artifacts, and Clinical Utility

Generative models, including Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), can introduce artifacts or inadequately capture clinically significant features, potentially limiting their usefulness in practical diagnostic contexts [32,63]. Rigorous expert validation is essential to evaluate synthetic images and narratives for their realism and fidelity to actual disease presentations. Furthermore, synthetic multimodal datasets must maintain coherent relationships between different data types (e.g., imaging, genomic, and textual clinical data) to support accurate and reliable diagnostic models [56,65].

7.3. Ethical and Regulatory Considerations

Ethical concerns regarding patient consent, data ownership, and potential misuse of synthetic datasets require clear communication and alignment with patient advocacy groups, particularly within rare disease communities [56]. Regulatory frameworks around synthetic data remain ambiguous, necessitating clear guidelines and policies that can ensure responsible use of synthetic datasets without compromising patient confidentiality or ethical standards. Navigating international regulatory barriers like GDPR and HIPAA requires synthetic datasets to maintain privacy and compliance without sacrificing data utility and clinical applicability [29,75].

7.4. Model Bias and Representativeness

Synthetic datasets risk propagating or amplifying existing biases inherent in original training data, potentially leading to skewed performance in clinical applications. Ensuring diverse, representative datasets is critical to preventing algorithmic biases that disproportionately affect certain patient populations, a particularly acute concern for rare diseases that often manifest differently across demographic groups [43,56]. Strategies to detect, mitigate, and continuously monitor bias within synthetic datasets are essential to ensure equitable clinical outcomes.

7.5. Enabling International Research Collaboration

Synthetic multimodal data generation significantly facilitates international research collaboration in rare disease research by addressing critical data scarcity, privacy concerns, and regulatory complexities. Synthetic datasets overcome cross-border regulatory hurdles by enabling researchers to share realistic but privacy-preserving data representations, facilitating multinational research efforts without compromising patient confidentiality [40,56]. By augmenting limited datasets, synthetic data enhances international data availability, supporting collaborative analyses and AI model development in rare disease diagnostics. Standardized synthetic data formats further streamline interoperability across diverse research institutions, significantly reducing technical barriers to collaboration and harmonizing international analytical approaches.

7.6. Practical Applications and Case Studies

Several practical implementations and case studies illustrate the effective use of synthetic multimodal data across clinical research settings (Table **??**). The MIMIC series, while focused on critical care, provides a model for synthetic data use in international collaboration, demonstrating how de-identified and synthetic datasets can substantially advance research capabilities while ensuring privacy [44]. Alzheimer's disease research has leveraged synthetic neuroimaging and clinical narratives to enhance early detection and progression monitoring, methodologies readily applicable to rare neurological disorders characterized by limited available data [36,65]. Similarly, synthetic genomic and phenotypic datasets support international research efforts into rare genetic disorders by enabling robust, privacy-preserving analysis and model development without direct patient data exposure [43,74].

7.7. Comparative Evaluation of Synthetic Data Generation Methods

Comparative analysis reveals that synthetic data generation methods differ significantly in their data types, clinical applications, advantages, and inherent limitations (Table **??**). GANs and VAEs excel in creating realistic multimodal datasets, enhancing diagnostic capabilities and supporting data augmentation for rare conditions, though they require careful validation and computational resources. Differential privacy and federated learning methods provide essential privacy protection and regulatory alignment, facilitating international research collaboration, but present challenges in complexity and model synchronization.

7.8. Future Directions and Research Priorities

Future research should focus on enhancing synthetic data realism, improving methods for rigorous clinical validation, and addressing ethical, regulatory, and interpretability challenges. Development of advanced explainable AI techniques to validate synthetic datasets and continuous monitoring for biases and data quality issues are critical. Collaborative international research efforts, standardized data generation protocols, and clear ethical and regulatory frameworks will further promote responsible and effective use of synthetic multimodal data in rare disease diagnostics and international healthcare research [29,43,75].

In conclusion, synthetic multimodal medical data provides a transformative opportunity for advancing diagnostics and research in rare diseases through international collaboration. Addressing data quality, realism, ethical concerns, and regulatory compliance remains crucial for maximizing the clinical utility and impact of these innovative methodologies.

8. Clinical Validation, Human-AI Collaboration, and Ethical Implications of Multimodal Generative AI

8.1. Comparative Validation of Multimodal AI and Human Clinicians

The advancement of multimodal generative AI in diagnostics necessitates rigorous comparative validation against human clinicians to ensure reliable clinical deployment. Comparative studies typically begin with meticulously curated multimodal datasets incorporating diverse data sources—imaging (radiological, pathological), genomic, and clinical notes. Trained multimodal AI models, such as transformer-based architectures and generative adversarial networks (GANs), are benchmarked against expert clinician panels through defined diagnostic scenarios, employing metrics such as accuracy, sensitivity, specificity, and area under the receiver operating characteristic curve (AUC-ROC) [24,27,39]. Notable studies illustrate AI's potential to match or exceed human diagnostic accuracy, particularly in oncology and neurology. For instance, multimodal AI systems integrating mammography with clinical data improved breast cancer detection accuracy from 83.6% to 90.6%, surpassing single-modality approaches [24]. Similarly, AI models combining MRI and PET scans significantly enhanced Alzheimer's disease diagnostic accuracy, demonstrating robust diagnostic performance in complex clinical scenarios [27]. Pathology-focused models, notably PathChat, have achieved diagnostic accuracies of up to 89.5%, clearly demonstrating advantages over single-modality and non-specialized AI systems [33].

However, these models face critical limitations identified through rigorous failure analyses, including sensitivity to variations in data quality, limited interpretability of complex diagnostic reasoning processes, challenges in capturing subtle contextual nuances, and the risk of generating plausible yet erroneous diagnostic outputs—commonly termed "hallucinations" [33,39]. Bias amplification in training data poses additional ethical challenges, potentially reinforcing disparities in healthcare outcomes across patient populations [56].

8.2. Human-AI Collaboration in Clinical Workflows

Effective clinical integration of multimodal generative AI emphasizes collaborative rather than autonomous roles, where AI systems augment rather than replace clinician expertise. The concept of trust calibration is critical, involving ensuring clinicians maintain appropriate confidence levels in AI recommendations through transparency and interpretability strategies. Techniques such as feature attribution, attention visualization, uncertainty quantification, and interactive visualization enable clinicians to understand AI reasoning, crucial for clinical acceptance [27,38]. Clinical scenarios exemplify varied levels of AI-driven decision support. Routine diagnostic tasks, such as triage or preliminary radiological interpretations, may involve higher automation levels. Complex clinical decisions, however, typically require robust human oversight to manage uncertainty and ethical nuances effectively. AI-human hybrid workflows have shown particular promise, leveraging AI strengths in processing large multimodal datasets to identify subtle clinical patterns while clinicians provide essential contextual judgment and ethical oversight [29,56].

8.3. Ethical, Regulatory, and Liability Considerations

As multimodal generative AI becomes integral to diagnostics, ethical and regulatory considerations become increasingly crucial. Responsibility and liability for AI-driven decisions necessitate clear role delineation among clinicians, developers, and regulators. Ethical challenges include ensuring informed consent, transparency of AI-generated outputs, and addressing potential data privacy risks arising from integrating sensitive multimodal patient data [56,58]. Fairness and equity considerations are paramount due to potential biases inherent in training datasets, risking exacerbation of healthcare disparities. Efforts to mitigate biases include rigorous bias audits, inclusive dataset curation, and continuous monitoring of model outputs across diverse patient populations [29]. Transparency and interpretability of AI reasoning processes are critical for clinical adoption, requiring investment in explainable AI (XAI) techniques and standardized reporting frameworks to enhance clinician trust and facilitate regulatory approvals [58].

Regulatory frameworks, including FDA guidelines, EU AI Act, and the AI/ML Action Plan, emphasize robust clinical validation, transparency, data governance, and continuous post-market monitoring to ensure patient safety and efficacy of multimodal AI tools. Addressing regulatory uncertainty involves creating agile governance frameworks capable of adapting to technological innovations while maintaining strict oversight to ensure ethical implementation [58,76].

8.4. Summary of Key Findings and Path Forward

Multimodal generative AI has demonstrated significant diagnostic advantages, improving clinical accuracy, efficiency, and personalized patient care across radiology, pathology, dermatology, ophthalmology, and neurology. While notable successes include enhanced breast cancer detection, Alzheimer's disease diagnosis, and pathology diagnostics, substantial challenges remain regarding data integration, model interpretability, and regulatory compliance. Future research priorities should include developing standardized multimodal datasets, advancing federated learning and privacy-preserving collaboration, improving model interpretability and explainability methods, and conducting rigorous clinical validation studies. Interdisciplinary collaboration among clinicians, researchers, regulators, and patient advocates is essential for addressing ethical concerns and regulatory challenges, ensuring responsible implementation and equitable healthcare delivery.

In conclusion, multimodal generative AI holds transformative potential for diagnostics, yet successful clinical integration demands careful attention to data quality, ethical considerations, interpretability, regulatory alignment, and human-AI collaboration. Addressing these multidimensional challenges through targeted research, policy development, and ethical governance will be crucial for leveraging AI's full potential in enhancing patient outcomes and advancing medical diagnostics.

9. Final Synthesis, Clinical Validation, Human-AI Collaboration, and Ethical Outlook for Multimodal Generative AI in Clinical Diagnostics

9.1. Current Status and Cross-specialty Clinical Impact

Multimodal generative AI has emerged as a transformative innovation in medical diagnostics by integrating heterogeneous data streams—including medical imaging, genomic profiles, clinical narratives, and electronic health records—into unified analytic frameworks that enhance diagnostic accuracy, clinical decision-making, and personalized care. Substantial advancements have been demonstrated across multiple specialties. In **radiology**, multimodal AI models that combine MRI, CT, PET, and clinical text have achieved exceptional performance in tasks such as automated report generation, disease classification, and abnormality detection, with large-scale benchmarks like MIMIC-CXR and ImageCLEF validating their clinical utility [41,44]. In **pathology**, diagnostic precision and prognostic capabilities have improved through the integration of histopathological images, genomic data, and clinical information. Notably, AI systems such as PathChat achieve diagnostic accuracy levels comparable to or exceeding expert pathologists, particularly in multimodal diagnostic and histopathology quality assurance tasks [33,60]. Meanwhile, **dermatology and ophthalmology** have rapidly embraced multimodal AI for early detection and diagnostic accuracy, leveraging dermoscopic images, OCT scans, patient histories, and synthetic data augmentation. Generative models in these fields support diagnostic training, rare disease recognition, and tailored treatment recommendations, while also enhancing patient education and workflow efficiency [34–37]. Despite this progress, full realization of multimodal AI's clinical potential remains constrained by critical challenges that demand targeted research and strategic solutions.

9.2. Clinical Validation and Human-AI Comparative Studies

A central requirement for clinical adoption is rigorous validation of multimodal AI performance against human clinicians. Comparative studies show that multimodal AI can match or even exceed human-level diagnostic capability in specific domains, particularly in breast cancer detection, Alzheimer's disease diagnosis, and rare pathology, where integration of imaging, text, and clinical data improves diagnostic accuracy beyond unimodal baselines [24,27,33]. Nonetheless, failure analyses highlight limitations, including sensitivity to data quality, risk of generating hallucinated or plausible yet erroneous outputs, and inability to fully capture subtle contextual nuances that clinicians often interpret intuitively [33,39]. Prospective, large-scale validation studies remain essential to confirm the real-world reliability of these systems. Such trials must include diverse patient populations, standardized protocols, and strategies for mitigating bias and improving interpretability, thereby strengthening clinician trust and regulatory alignment [33,58].

Table 9. Comparative validation of multimodal generative AI vs. human clinicians

Clinical Task	**Multimodal AI Model**	**Performance Outcomes**	**Identified Limitations and Insights**
Breast Cancer Detection / Imaging	Multimodal imaging + clinical data models [24]	Accuracy: 90.6%, surpassing human radiologists	Data sensitivity, dependency on data quality, interpretability challenges
Alzheimer's Disease Diagnosis / Neurological Diagnostics	MLG-GAN + Multimodal Transformer [27]	Superior to state-of-the-art imaging models for early detection	Model complexity, limited interpretability, sensitivity to data quality
Pathology Diagnostics	PathChat Multimodal Model [33]	Accuracy up to 89.5%, exceeding unimodal methods	Limited generalizability, hallucination risks, interpretability concerns
Clinical Oncology	Multimodal chatbot evaluations [39]	Comparable or superior to expert clinicians in certain diagnostic tasks	Variability across scenarios, interpretability issues, generalization limits

9.3. Human-AI Collaboration Models and Trust Calibration

The most effective paradigm for clinical deployment is not full automation but hybrid human-AI collaboration, in which AI augments clinician decision-making rather than replacing expertise. Central to this approach is *trust calibration*, ensuring clinicians neither over- nor under-rely on AI recommendations. Techniques such as attention visualization, feature attribution, uncertainty quantification, and interactive interfaces allow clinicians to interrogate AI-generated outputs and align them with clinical reasoning [27,38]. These strategies promote transparency, foster trust, and ultimately improve efficiency, accuracy, and outcomes by combining the complementary strengths of human judgment and machine precision. For example, in breast cancer screening, multimodal AI can provide preliminary assessments supported by attribution maps, while clinicians make final determinations; in neurology, probability assessments can flag early disease progression while preserving human oversight; and in critical care, risk stratification models can optimize ICU resource allocation while clinicians address ethical and contextual nuances. Case studies summarized in Table 10 illustrate how automation levels and trust calibration techniques influence outcomes across domains ranging from oncology to psychiatry.

Table 10. Human-AI Collaboration in Clinical Diagnostics

Application	Automation Level	Trust Calibration Approaches	Clinical Impact and Issues
Breast Cancer Detection	Decision support (preliminary assessments)	Feature attribution, uncertainty quantification	Improved detection rate, risk of clinician over-reliance [9]
Neurological Disorders	Moderate automation (probability assessments)	Attention visualization, validation studies	Early detection, interpretability concerns [27]
ICU Risk Management	Risk stratification	Confidence intervals, interactive visualizations	Resource optimization, potential biases [29]
Psychiatric Evaluation	Hybrid decision-making	Interactive interfaces, uncertainty modeling	Improved personalized care, privacy challenges [56]

9.4. Ethical, Regulatory, and Liability Challenges

The integration of multimodal generative AI into diagnostics raises substantial ethical, regulatory, and liability concerns that must be systematically addressed to ensure safe and equitable use. Responsibility frameworks must clearly delineate roles and accountability among clinicians, developers, and regulators, particularly in the event of diagnostic errors or AI-generated inaccuracies [58]. Issues of fairness and equity are paramount, as algorithmic biases introduced by non-representative training data can reinforce healthcare disparities; solutions include proactive bias audits, curated diverse datasets, and continuous monitoring across demographic groups [29]. Transparency and interpretability of model reasoning processes remain critical for clinician trust and patient safety, necessitating further development of explainable AI methods and standardized reporting protocols [58]. Meanwhile, regulatory frameworks must adapt to international variability in oversight; agile and adaptive guidelines are required to balance innovation with patient safety, particularly as multimodal systems evolve rapidly in capability and complexity [75]. Table 11 summarizes these challenges, proposed solutions, and persistent uncertainties.

Table 11. Ethical and Regulatory Considerations of Multimodal Generative AI

Aspect	Issues	Proposed Solutions	Ongoing Challenges
Responsibility	Liability, clinician-AI delineation	Role definition, accountability frameworks	Regulatory uncertainty, liability clarification [58]
Fairness/Equity	Data biases, healthcare disparities	Diverse datasets, regular audits	Persistent dataset representation challenges [29]
Transparency	Model interpretability, clinical trust	Explainable AI techniques, standardized protocols	Performance vs. interpretability trade-off [58]
Regulatory Compliance	International regulation disparities	Agile, adaptive regulatory guidelines	Balancing innovation with oversight [75]

9.5. Future Clinical, Research, and Regulatory Outlook

Realizing the full potential of multimodal generative AI requires sustained efforts in research, clinical validation, and governance. Future priorities include the development of large-scale, diverse, and standardized multimodal datasets to enable representative training and validation; privacy-preserving approaches such as federated learning to facilitate global collaboration while safeguarding confidentiality; and advancement of interpretability techniques that strengthen clinician trust and regulatory compliance. Equally critical are rigorous, prospective multi-site clinical trials designed to evaluate real-world performance across varied healthcare settings and patient populations. Progress also depends on close collaboration among clinicians, AI researchers, regulators, and patient advocacy groups, ensuring that models are designed, validated, and deployed in ways that reflect both technical rigor and ethical responsibility.

- Development of large-scale, diverse, and standardized multimodal datasets for comprehensive training and validation.
- Advancement of privacy-preserving federated learning approaches to facilitate international collaboration without compromising patient confidentiality.
- Refinement of AI interpretability and explainability techniques to build clinical trust and regulatory compliance.
- Implementation of rigorous, prospective clinical trials for validation of multimodal generative AI systems in diverse real-world clinical settings.
- Multidisciplinary collaboration between clinicians, researchers, regulators, and patient advocacy groups to ensure responsible and ethical AI development.

By addressing these areas, multimodal generative AI can achieve its promise of improving diagnostic accuracy, clinical efficiency, and personalized care, while ensuring equitable outcomes across populations. Ultimately, proactive governance, continuous monitoring, and adaptive regulatory frameworks will be required to align innovation with patient safety. Figure 8 presents a predictive roadmap for the field, highlighting milestones such as the establishment of standardized multimodal datasets in the near term, the progression to prospective multi-site trials, the implementation of adaptive regulatory frameworks, and eventual widespread integration into clinical practice with continuous post-market monitoring.

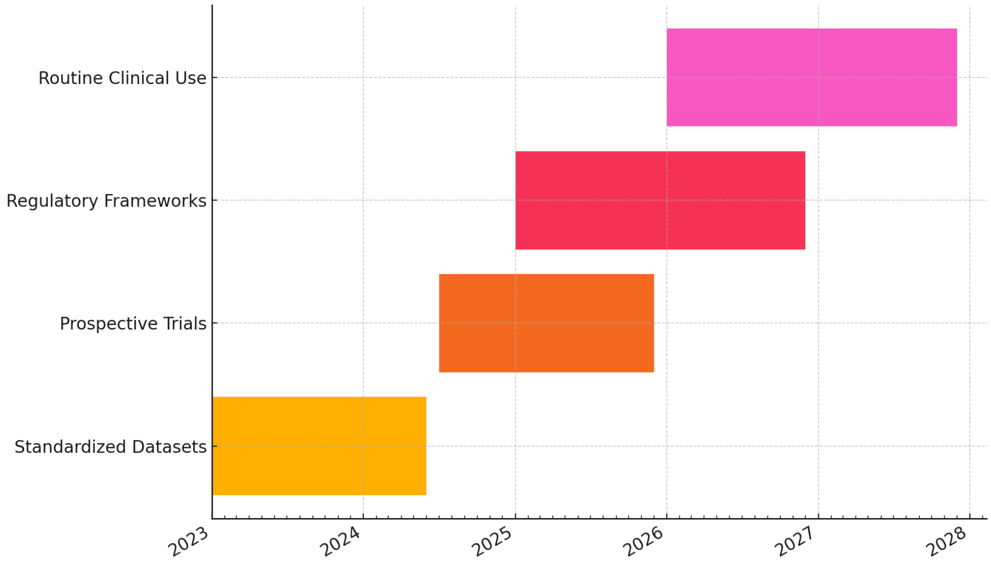


Figure 8. Predictive roadmap of field milestones—from standardized multimodal datasets and benchmarking, to prospective multi-site clinical trials and adaptive regulatory frameworks (CONSORT-AI, SPIRIT-AI, DECIDE-AI, TRIPOD+AI), culminating in routine clinical use with continuous post-market monitoring and quality improvement [18–23].

10. Conclusion and Strategic Priorities for Multimodal Generative AI in Clinical Diagnostics

Multimodal generative artificial intelligence (AI) represents a transformative opportunity for clinical diagnostics, holding immense promise to significantly enhance patient care, diagnostic accuracy, and healthcare efficiency. Realizing this potential, however, necessitates addressing critical clinical, research, ethical, and regulatory priorities. This call to action highlights the essential areas of strategic focus required for responsible adoption and optimal utilization of multimodal generative AI technologies in clinical practice.

10.1. High-Priority Clinical Domains

Immediate clinical priorities span diverse medical fields where multimodal AI integration could substantially enhance diagnostic capabilities and patient outcomes. Oncology stands out as a prime area, where integration of imaging, pathology, genomic profiles, and clinical histories offers the potential to revolutionize early cancer detection, precise characterization, personalized treatment strategies, and continuous therapeutic monitoring [38]. Neurology, cardiology, rare disease diagnostics, and emergency medicine also represent priority areas where multimodal data integration can significantly enhance diagnostic accuracy, timely interventions, and patient management through comprehensive analyses of imaging, electrophysiological, clinical, and genomic data [27,33,40].

10.2. Strategic Research Priorities

Advancing multimodal generative AI necessitates targeted research in several critical domains:

- **Development of Advanced Model Architectures**: Emphasis on novel deep learning methods (e.g., attention mechanisms, transformers, graph neural networks) optimized for integrating complex, heterogeneous medical datasets [38].
- **Large-scale Dataset Creation and Standardization**: Collaborative efforts are required to build robust, diverse, and representative multimodal medical datasets, addressing data scarcity, bias, and standardization challenges crucial for model training and validation [28,40].
- **Enhanced Model Interpretability and Transparency**: Further investment in explainable AI (XAI) techniques will ensure clinicians understand AI-generated outputs, fostering trust and enabling responsible clinical integration [29,33].
- **Rigorous Clinical Validation**: Conducting prospective clinical trials to rigorously assess real-world efficacy, generalizability, and safety of multimodal generative AI systems in diverse health-care settings [29,75].
- **Privacy-preserving and Federated Learning Approaches**: Continued research into methods that facilitate international collaboration without compromising patient confidentiality, thus overcoming regulatory barriers and enhancing global AI research cooperation [29,40].
- **Data Fusion and Uncertainty Quantification**: Developing robust multimodal data fusion methods that effectively handle missing, incomplete, or heterogeneous data, combined with accurate uncertainty quantification to assist clinical decision-making [33,38].
- **Continual Learning Frameworks**: Ensuring that multimodal AI models can dynamically adapt to new data and evolving clinical practices, maintaining long-term performance and relevance in changing clinical environments.

10.3. Ethical and Regulatory Frameworks

Implementing multimodal generative AI requires rigorous ethical and regulatory frameworks addressing responsibility, fairness, transparency, and interpretability. Clear delineation of roles among clinicians, developers, and regulators is critical for addressing liability concerns associated with AI-driven diagnostics [58]. Ensuring fairness and equity involves proactive monitoring and mitigation of biases inherent in datasets and algorithms, promoting equitable healthcare access and outcomes [29]. Transparency in AI decision-making processes, supported by robust XAI methodologies, is essential to building clinician trust and facilitating regulatory compliance. Internationally harmonized regulatory guidelines should provide agile oversight frameworks that balance innovation with patient safety, addressing both current technological capabilities and future advancements [29,58].

10.4. Interdisciplinary Collaboration and Stakeholder Engagement

Realizing the potential of multimodal generative AI requires interdisciplinary collaboration involving clinical researchers, AI specialists, data scientists, biomedical informaticians, ethicists, regulators, healthcare administrators, and patient advocacy groups. This diverse collaboration fosters comprehensive solutions to technical, ethical, and operational challenges. Active engagement of patients and advocacy communities ensures that AI tools align with patient-centered priorities and ethical standards, promoting transparency, consent, and equitable healthcare delivery.

10.5. Strategic Actions for Overcoming Barriers and Accelerating Adoption

To accelerate responsible integration and adoption of multimodal generative AI, several strategic actions are recommended:

- **Establish Clear Regulatory Pathways**: Collaborate with regulatory bodies to develop explicit validation and approval frameworks, streamlining the translation of AI technologies from research to clinical practice.
- **Invest in Computational Infrastructure**: Support healthcare institutions in building robust computational platforms capable of integrating and deploying complex multimodal AI models efficiently.
- **Implement Comprehensive Training Programs**: Educate clinicians and healthcare professionals in effectively using AI-assisted diagnostic tools, emphasizing practical application, interpretability, and ethical considerations.
- **Promote Open Science and Data Sharing Initiatives**: Encourage international sharing of algorithms, models, and standardized datasets to foster reproducible, collaborative, and transparent AI research.
- **Conduct Cost-effectiveness and Impact Studies**: Demonstrate the economic and clinical value of multimodal AI to drive reimbursement and justify widespread clinical implementation.
- **Develop Standardized Best Practices and Ethical Guidelines**: Establish international guidelines and best practices for the responsible development, validation, and clinical deployment of multimodal AI systems.
- **Engage Public and Community Dialogues**: Foster ongoing communication with patients and communities to build trust, address concerns, and ensure transparency around AI usage in healthcare settings.

10.6. Summary of Strategic Priorities and Impact Outlook

A structured summary of strategic priorities, stakeholders, and recommended actions for advancing multimodal generative AI in clinical diagnostics is presented in Table 12, highlighting immediate clinical priorities, focused research initiatives, key collaborative stakeholders, and targeted strategic actions essential for successful adoption and impact.

Table 12. Strategic Summary of Clinical Priorities, Research Directions, Stakeholders, and Recommended Actions

Clinical Priorities	Research Directions	Collaborative Stakeholders	Strategic Actions
Oncology	Dataset creation, multimodal fusion, interpretability	Clinical researchers, AI specialists, Data scientists, Regulators	Regulatory framework development, clinician training, standardized data sharing protocols
Neurology	Advanced model architectures, clinical validation, federated learning	Neurologists, AI researchers, Ethicists, Biomedical informaticians	Infrastructure investment, prospective clinical trials, privacy-preserving collaboration
Cardiology	Multimodal integration, uncertainty quantification, continual learning	Cardiologists, Data scientists, Healthcare administrators, Regulatory bodies	Robust model validation, clinician education, continuous performance monitoring
Rare diseases	Synthetic data generation, federated learning, bias mitigation	Geneticists, Patient advocates, Regulators, Ethicists	International collaboration, open science initiatives, ethical governance frameworks
Emergency medicine	Rapid multimodal fusion, interpretability, infrastructure enhancement	Emergency physicians, Clinical researchers, AI specialists, Healthcare administrators	Real-time model integration, clinician decision-support training, ethical oversight mechanisms

Multimodal generative AI holds extraordinary promise for reshaping clinical diagnostics, offering substantial improvements in diagnostic accuracy, personalized patient care, clinical efficiency, and health system optimization. Yet, realizing this potential demands focused action across clinical, research, ethical, and regulatory domains. Through interdisciplinary collaboration, strategic investment, comprehensive validation, proactive ethical governance, and international standardization efforts, multimodal generative AI can be harnessed responsibly to significantly enhance patient outcomes and healthcare delivery, ultimately advancing the goal of precision medicine in clinical diagnostics.

Author Contributions: Conceptualization, M.M.; methodology, M.M.; software, M.M.; validation, M.M.; formal analysis, M.M.; investigation, M.M.; resources, S.A.G.; data curation, M.M.; writing—original draft preparation, M.M.; writing—review and editing, M.M. and S.A.G.; supervision, M.M.; project administration, M.M. and S.A.G.; All authors have read and agreed to the published version of the manuscript.

Funding: No funding was obtained for this research.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: All the references used in this research review have been obtained from publicly available PubMed research repository.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
LLM	Large Language Model
VQA	Visual Question Answering
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
CT	Computed Tomography
XAI	Explainable Artificial Intelligence
Grad-CAM	Gradient-weighted Class Activation Mapping
SHAP	SHapley Additive exPlanations
LIME	Local Interpretable Model-agnostic Explanations
Med-PaLM	Medical Pathways Language Model
LLaVA-Med	Large Language and Vision Assistant for Medicine
BiomedGPT	Biomedical Generative Pre-trained Transformer
BioGPT-ViT	BioGPT Vision Transformer
MLG-GAN	Multi-Level Guided Generative Adversarial Network
Mul-T	Multimodal Transformer
ECG	Electrocardiogram
EEG	Electroencephalogram
FDG-PET	Fluorodeoxyglucose Positron Emission Tomography
GANs	Generative Adversarial Networks
ROCO	Radiology Objects in COntext
NIH14	National Institutes of Health Chest X-ray Dataset
QA	Question Answering
OCT	Optical Coherence Tomography
MIMIC	Medical Information Mart for Intensive Care
CLARO	CT Imaging of Lung Cancer And Related Outcomes
LCID	Lung Cancer Imaging Database
EHR	Electronic Health Records
AUC-ROC	Area Under the Receiver Operating Characteristic curve

BLEU	Bilingual Evaluation Understudy
MIMIC-CXR	Medical Information Mart for Intensive Care Chest X-Ray
ImageCLEF	Image Cross-Language Evaluation Forum
MMBERT	Multi-Modal Bidirectional Encoder Representations from Transformers
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
CONCH	Contrastive Learning from Captions for Histopathology
IHC	Immunohistochemistry
H&E	Hematoxylin and Eosin
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
Onto-CGAN	Ontology-enhanced Conditional Generative Adversarial Network
AML	Acute Myeloid Leukemia
TPLC	Total Product Life Cycle
PMA	Premarket Approval
SaMD	Software as a Medical Device
FDA	Food and Drug Administration

References

1. Shahparvari, S.; Hassanizadeh, B.; Mohammadi, A.; Kiani, B.; Lau, K.H.; Chhetri, P.; Abbasi, B. A decision support system for prioritised COVID-19 two-dosage vaccination allocation and distribution. *Transportation Research Part E: Logistics and Transportation Review* **2022**, *159*, 102598.
2. Maleki, M.; Khan, M. Covid-19 health equity & justice dashboard: A step towards countering health disparities among seniors and minority population. *Available at SSRN 4595845* **2023**.
3. Maleki, M. Advancing Healthcare Accessibility through a Neighborhood Search Recommendation Tool. *Available at SSRN 4825773* **2024**.
4. Maleki, M.; Ghahari, S. Clinical trials protocol authoring using llms. *arXiv preprint arXiv:2404.05044* **2024**.
5. Rieke, N.; Hancox, J.; Li, W.; Milletarì, F.; Roth, H.R.; Albarqouni, S.; Bakas, S.; Galtier, M.N.; Landman, B.A.; Maier-Hein, K.; et al. The future of digital health with federated learning. *NPJ Digital Medicine* **2020**, *3*, 119. https://doi.org/10.1038/s41746-020-00323-1.
6. Su, Z.; Liang, B.; Shi, F.; Gelfond, J.; Šegalo, S.; Wang, J.; Jia, P.; Hao, X. Deep learning-based facial image analysis in medical research: a systematic review protocol. *BMJ Open* **2021**, *11*, e047549. https://doi.org/10.1136/bmjopen-2020-047549.
7. Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* **2019**, *25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7.
8. Deo, R.C. Machine Learning in Medicine. *Circulation* **2015**, *132*, 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593.
9. Liu, X.; Faes, L.; Kale, A.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdas, M.; Kern, C.; et al. Artificial intelligence in clinical medicine: A review of its foundations, applications, and future directions. *NPJ Digit Med* **2023**, *6*, 15. https://doi.org/10.1038/s41746-023-00790-6.
10. Maleki, M. Clustering analysis of us covid-19 rates, vaccine participation, and socioeconomic factors. *arXiv preprint arXiv:2404.08186* **2024**.
11. Shahparvari, S.; Hassanizadeh, B.; Chowdhury, P.; Lau, K.H.; Chhetri, P.; Childerhouse, P. Supply Chain Strategies to Reduce Vaccine Wastage for Disease X: A Covid-19 Case. *Available at SSRN 5092993* **2024**.
12. Maleki, M.; Bahrami, M.; Menendez, M.; Balsa-Barreiro, J. Social Behavior and COVID-19: Analysis of the Social Factors behind Compliance with Interventions across the United States. *International Journal of Environmental Research and Public Health* **2022**, *19*.
13. Maleki, M.; Ghahari, S. Comprehensive clustering analysis and profiling of covid-19 vaccine hesitancy and related factors across us counties: Insights for future pandemic responses. In Proceedings of the Healthcare. MDPI, 2024, Vol. 12, p. 1458.
14. Esteva, A.; Robicquet, A.; Ramsundar, B.; Kuleshov, V.; DePristo, M.; Chou, K.; Cui, C.; Corrado, G.; Thrun, S.; Dean, J. A guide to deep learning in healthcare. *Nature Medicine* **2019**, *25*, 24–29. https://doi.org/10.1038/s41591-018-0316-z.
15. Maleki, M.; Ghahari, S. Impact of Major Health Events on Pharmaceutical Stocks: A Comprehensive Analysis Using Macroeconomic and Market Indicators. *arXiv preprint arXiv:2408.01883* **2024**.

16. Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* **2019**, *25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7.
17. Haug, C.J.; Drazen, J.M. Artificial Intelligence and Machine Learning in Clinical Medicine, 2023. *The New England Journal of Medicine* **2023**, *388*, 1201–1208. https://doi.org/10.1056/NEJMra2302038.
18. Liu, X.; Rivera, S.C.; Moher, D.; Calvert, M.J.; Denniston, A.K.; SPIRIT-AI.; Group, C.A.W. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. *BMJ* **2020**, *370*, m3164. https://doi.org/10.1136/bmj.m3164.
19. Rivera, S.C.; Liu, X.; Chan, A.W.; Denniston, A.K.; Calvert, M.J.; SPIRIT-AI.; Group, C.A.W. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI Extension. *BMJ* **2020**, *370*, m3210. https://doi.org/10.1136/bmj.m3210.
20. Vasey, B.; Nagendran, M.; Campbell, B.; Clifton, D.A.; Collins, G.S.; Denaxas, S.; Denniston, A.K.; Faes, L.; Geerts, H.; Ibrahim, M.; et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ* **2022**, *377*, e070904. https://doi.org/10.1136/bmj-2022-070904.
21. Collins, G.S.; Dhiman, P.; Andaur Navarro, C.L.; Keogh, R.H.; van Smeden, M.; Riley, R.D.; Wolff, R.; Damen, J.A.; Verbakel, J.; et al.. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378. https://doi.org/10.1136/bmj-2023-078378.
22. Feng, J.; Phillips, R.V.; Malenica, I.; Bishara, A.; Hubbard, A.E.; Celi, L.A.; Pirracchio, R. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. *NPJ Digital Medicine* **2022**, *5*, 66. https://doi.org/10.1038/s41746-022-00611-y.
23. Matheny, M.E.; Whicher, D.; Thadaney Israni, S. Artificial Intelligence in Health Care: A Report From the National Academy of Medicine. *JAMA* **2020**, *323*, 509–510. https://doi.org/10.1001/jama.2019.21579.
24. F, A.; Y, A.; S, A.M.; A, B.; IM, T.; R, H. Histopathology in focus: a review on explainable multi-modal approaches for breast cancer diagnosis. *Frontiers in medicine* **2024**, *11*, 1450103.
25. EK, H.; J, H.; B, R.; J, G.; B, P.; S, K.; K, Y.; J, E.; B, B.; JB, J.; et al. Diagnostic Accuracy and Clinical Value of a Domain-specific Multimodal Generative AI Model for Chest Radiograph Report Generation. *Radiology* **2025**, *314*, e241476. https://doi.org/10.1148/radiol.241476.
26. Mittal, S.; Tong, A.; Young, S.W.; Jha, P. Artificial intelligence applications in endometriosis imaging. *Abdominal Radiology* **2025**. Online ahead of print, https://doi.org/10.1007/s00261-025-04897-w.
27. X, G.; F, S.; D, S.; M, L. Multimodal transformer network for incomplete image generation and diagnosis of Alzheimer's disease. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* **2023**, *110*, 102303. https://doi.org/S0895-6111(23)00121-0[pii]10.1016/j.compmedimag.2023.102303.
28. E, U.; MM, B.; A, W.; G, R.; R, S.; A, S.; H, G.; AV, P. Multimodal Generative AI for Anatomic Pathology-A Review of Current Applications to Envisage the Future Direction. *Advances in anatomic pathology* **2025**. https://doi.org/10.1097/PAP.0000000000000498.
29. S, R. Generative AI in healthcare: an implementation science informed translational path on application, integration and governance. *Implementation science : IS* **2024**, *19*, 27. https://doi.org/10.1186/s13012-024-01357-927.
30. A, A. CareAssist GPT improves patient user experience with a patient centered approach to computer aided diagnosis. *Scientific reports* **2025**, *15*, 22727. https://doi.org/10.1038/s41598-025-01518-w22727.
31. GS, H.; M, J.; S, K.; K, C.; J, J.; GY, L.; K, S.; KD, K.; SM, R.; JB, S.; et al. Overcoming the Challenges in the Development and Implementation of Artificial Intelligence in Radiology: A Comprehensive Review of Solutions Beyond Supervised Learning. *Korean journal of radiology* **2023**, *24*, 1061–1080. https://doi.org/10.3348/kjr.2023.0393.
32. N, P.; SW, J.; JH, J.; TK, M. Multimodal AI in Biomedicine: Pioneering the Future of Biomaterials, Diagnostics, and Personalized Healthcare. *Nanomaterials (Basel, Switzerland)* **2025**, *15*. https://doi.org/10.3390/nano15120895895.
33. Y, G.; P, W.; Y, L.; Y, S.; H, Q.; X, Z.; H, P.; Y, G.; C, L.; Z, G.; et al. Application of artificial intelligence in the diagnosis of malignant digestive tract tumors: focusing on opportunities and challenges in endoscopy and pathology. *Journal of translational medicine* **2025**, *23*, 412. https://doi.org/10.1186/s12967-025-06428-z412.
34. Z, K.; M, A.; B, J.; N, B.; J, S. The Role of ChatGPT in Dermatology Diagnostics. *Diagnostics (Basel, Switzerland)* **2025**, *15*.

35. A, T.; ML, W.; A, H.; C, L.; M, S.; EY, L. Artificial intelligence-enabled precision medicine for inflammatory skin diseases. *ArXiv* **2025**.
36. SC, S.; M, S.; F, A.; J, H.; PA, K. Generative artificial intelligence in ophthalmology: current innovations, future applications and challenges. *The British journal of ophthalmology* **2024**, *108*, 1335–1340. https://doi.org/10.1136/bjo-2024-325458e325458.
37. JR, T.; J, D.; X, Z.; KW, L.; K, H.; X, W. Advancing healthcare through multimodal data fusion: a comprehensive review of techniques and applications. *PeerJ. Computer science* **2024**, *10*, e2298. https://doi.org/10.7717/peerj-cs.2298e2298.
38. J, L.; RJ, C.; B, C.; MY, L.; M, B.; D, S.; AJ, V.; C, C.; L, Z.; DFK, W.; et al. Artificial intelligence for multimodal data integration in oncology. *Cancer cell* **2022**, *40*, 1095–1110. https://doi.org/S1535-6108(22)00441-X[pii]10.1016/j.ccell.2022.09.012.
39. HH, R.; J, P.; A, C.; B, F.; Y, W.; RR, G.; A, T.; M, D.; S, A.; E, G.; et al. Generative Artificial Intelligence in Pathology and Medicine: A Deeper Dive. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **2025**, *38*, 100687. https://doi.org/S0893-3952(24)00267-9[pii]10.1016/j.modpat.2024.100687.
40. Yoo, J.; Kim, J.; Kim, B.; Jeong, B. Exploring characteristic features of attention-deficit/hyperactivity disorder: Findings from multi-modal MRI and candidate genetic data. *Brain Imaging and Behavior* **2020**, *14*, 2132–2147. https://doi.org/10.1007/s11682-019-00164-x.
41. D, B.; V, S.; Y, B.; E, K.; BS, G.; GN, N.; E, K. Assessing GPT-4 multimodal performance in radiological image analysis. *European radiology* **2025**, *35*, 1959–1965. https://doi.org/10.1007/s00330-024-11035-5.
42. VM, R.; M, H.; M, M.; S, A.; S, K.; EJ, T.; P, R. Multimodal generative AI for medical image interpretation. *Nature* **2025**, *639*, 888–896. https://doi.org/10.1038/s41586-025-08675-y.
43. M, I.; YA, K.; S, A.; C, S.; M, B.; J, P.; B, E.; G, E.; M, D. Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges. *Computers in biology and medicine* **2025**, *189*, 109834. https://doi.org/S0010-4825(25)00184-2[pii]10.1016/j.compbiomed.2025.109834.
44. IU, H.; M, M.; M, A.H.; H, O.; ZY, H.; Z, L. Advancements in Medical Radiology Through Multimodal Machine Learning: A Comprehensive Overview. *Bioengineering (Basel, Switzerland)* **2025**, *12*. https://doi.org/10.3390/bioengineering12050477477.
45. J, S.; J, M.; Q, Z.; W, L.; C, W. Predicting gene mutation status via artificial intelligence technologies based on multimodal integration (MMI) to advance precision oncology. *Seminars in cancer biology* **2023**, *91*, 1–15.
46. Acosta, J.N.; Falcone, G.J.; Rajpurkar, P.; Topol, E.J. Multimodal biomedical AI. *Nature Medicine* **2022**, *28*, 1773–1784. https://doi.org/10.1038/s41591-022-01981-2.
47. Stahlschmidt, S.R.; Ulfenborg, B.; Synnergren, J. Multimodal deep learning for biomedical data fusion: methods, applications, and challenges. *Briefings in Bioinformatics* **2022**, *23*, bbab569. https://doi.org/10.1093/bib/bbab569.
48. Kline, A.; Cagnazzo, F.; Finlayson, S.G.; Beam, A.L.; Celi, L.A.; Szolovits, P.; Naumann, T.; Chen, I.Y. Multimodal machine learning in precision health: a scoping review. *NPJ Digital Medicine* **2022**, *5*, 171. https://doi.org/10.1038/s41746-022-00712-8.
49. Lambert, B.; Forbes, F.; Doyle, S.; Dehaene, H.; Dojat, M. A unified review of uncertainty quantification in deep learning models for medical image analysis. *Artificial Intelligence in Medicine* **2024**, *150*, 102830. https://doi.org/10.1016/j.artmed.2024.102830.
50. P, R.; B, K.; S, F.; M, M.; MM, S.; P, R.; B, E. A Current Review of Generative AI in Medicine: Core Concepts, Applications, and Current Limitations. *Current reviews in musculoskeletal medicine* **2025**, *18*, 246–266. https://doi.org/10.1007/s12178-025-09961-y.
51. Jabbour, S.; Fouhey, D.; Kazerooni, E.; Wiens, J.; Sjoding, M.W. Combining chest X-rays and electronic health record (EHR) data using machine learning to diagnose acute respiratory failure. *Journal of the American Medical Informatics Association* **2022**, *29*, 1060–1068. https://doi.org/10.1093/jamia/ocac030.
52. Huang, S.C.; Pareek, A.; Jensen, M.; Lungren, M.P.; Chaudhari, A.S.; Yeung, S.; Ng, A.Y.; Lungren, M.P. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Scientific Reports* **2020**, *10*, 20597. https://doi.org/10.1038/s41598-020-78888-w.
53. Cahan, N.; Klang, E.; Marom, E.M.; Soffer, S.; Barash, Y.; Burshtein, E.; Konen, E.; Greenspan, H. Multimodal fusion models for pulmonary embolism mortality prediction. *Scientific Reports* **2023**, *13*, 7544. https://doi.org/10.1038/s41598-023-34303-8.

54. Fiorini, L.; Manani, G.; Tartarisco, G.; Cavallo, F.; Vissentin, M.; Cenci, A.; et al. Early Detection of Sepsis With Machine Learning Techniques: A Brief Clinical Perspective. *Frontiers in Medicine* **2021**, *8*, 617486. https://doi.org/10.3389/fmed.2021.617486.
55. Vairetti, G.; Giannaccare, G.; Pellegrini, M.; Vagge, A.; Bacherini, D.; Cagini, C.; .; et al. Deep learning methods for age-related macular degeneration across multimodal ophthalmic imaging: a systematic review and meta-analysis. *PLOS ONE* **2024**, *19*, e0299897. https://doi.org/10.1371/journal.pone.0299897.
56. L, T. Beyond Discrimination: Generative AI Applications and Ethical Challenges in Forensic Psychiatry. *Frontiers in psychiatry* **2024**, *15*, 1346059. https://doi.org/10.3389/fpsyt.2024.13460591346059.
57. KN, K. Generative Artificial Intelligence and Musculoskeletal Health Care. *HSS journal : the musculoskeletal journal of Hospital for Special Surgery* **2025**, p. 15563316251335334. https://doi.org/10.1177/1556331625133533415563316251335334.
58. SS, H.; MS, F.; JJ, W.; KN, K.; PN, R. Ethical Application of Generative Artificial Intelligence in Medicine. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* **2025**, *41*, 874–885.
59. C, C.; W, S.; Y, W.; Z, Z.; X, H.; Y, J. The path from task-specific to general purpose artificial intelligence for medical diagnostics: A bibliometric analysis. *Computers in biology and medicine* **2024**, *172*, 108258. https://doi.org/S0010-4825(24)00342-1[pii]10.1016/j.compbiomed.2024.108258.
60. MY, L.; B, C.; DFK, W.; RJ, C.; M, Z.; AK, C.; K, I.; A, K.; D, P.; A, P.; et al. A multimodal generative AI copilot for human pathology. *Nature* **2024**, *634*, 466–473. https://doi.org/10.1038/s41586-024-07618-3.
61. X, F.; E, S.; A, W. Application of digital pathology-based advanced analytics of tumour microenvironment organisation to predict prognosis and therapeutic response. *The Journal of pathology* **2023**, *260*, 578–591. https://doi.org/10.1002/path.6153.
62. G, V.M.; de S Santos R, L.; L, H.S.V.; de Paiva A, C.; de Alcantara Dos Santos Neto P. XRaySwinGen: Automatic medical reporting for X-ray exams with multimodal model. *Heliyon* **2024**, *10*, e27516. https://doi.org/10.1016/j.heliyon.2024.e27516e27516.
63. B, S.; DM, R.; G, R.; A, P. Evaluating the Clinical Realism of Synthetic Chest X-Rays Generated Using Progressively Growing GANs. *SN computer science* **2021**, *2*, 321. https://doi.org/10.1007/s42979-021-00720-7321.
64. SA, A.; ZH, K.; M, E.; M, S.; ARH, A.; MB, K.; T, T.; M, J.; AC, P.; H, H.; et al. Generative Adversarial Networks in Digital Histopathology: Current Applications, Limitations, Ethical Considerations, and Future Directions. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **2024**, *37*, 100369. https://doi.org/S0893-3952(23)00274-0[pii]10.1016/j.modpat.2023.100369.
65. J, W.; VH, K. Towards generative digital twins in biomedical research. *Computational and structural biotechnology journal* **2024**, *23*, 3481–3488. https://doi.org/10.1016/j.csbj.2024.09.030.
66. Skandarani, Y.; Jodoin, P.M.; Lalande, A. GANs for Medical Image Synthesis: An Empirical Study. *J Imaging* **2023**, *9*, 69. https://doi.org/10.3390/jimaging9030069.
67. Dorjsembe, Z.; Pao, H.K.; Odonchimed, S.; Xiao, F. Conditional Diffusion Models for Semantic 3D Brain MRI Synthesis. *IEEE J Biomed Health Inform* **2024**, *28*, 4084–4093. https://doi.org/10.1109/JBHI.2024.3385504.
68. Kidder, B.L. Advanced image generation for cancer using diffusion models. *Biol Methods Protoc* **2024**, *9*, bpae062. https://doi.org/10.1093/biomethods/bpae062.
69. Bradley, W.; Rockenschaub, P.; Guo, W.; Li'o, P. Synthetic electronic health records generated with variational graph autoencoders. *PLoS One* **2023**, *18*, e0283487. https://doi.org/10.1371/journal.pone.0283487.
70. Smolyak, D.; Bjarnad'ottir, M.V.; Crowley, K.; Agarwal, R. Large language models and synthetic health data: progress and prospects. *JAMIA Open* **2024**, *7*, ooae114. https://doi.org/10.1093/jamiaopen/ooae114.
71. Litake, O.; Park, B.H.; Tully, J.L.; Gabriel, R.A. Constructing synthetic datasets with generative artificial intelligence to train large language models to classify acute renal failure from clinical notes. *J Am Med Inform Assoc* **2024**, *31*, 1404–1410. https://doi.org/10.1093/jamia/ocae081.
72. Sheller, M.J.; Reina, G.A.; Edwards, B.; Martin, J.; Pati, S.; Kotrotsou, A.; Milchenko, M.; Xu, W.; Marcus, D.S.; Colen, R.R.; et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* **2020**, *10*, 12598. https://doi.org/10.1038/s41598-020-69250-1.
73. Ficek, J.; Wang, W.; Chen, H.; Dagne, G.; Daley, E. Differential privacy in health research: A scoping review. *J Am Med Inform Assoc* **2021**, *28*, 2269–2276. https://doi.org/10.1093/jamia/ocab135.
74. JN, E.; W, H.; C, R.; S, S.; U, P.; C, M.T.; H, S.; CD, B.; C, S.; K, S.E.; et al. Mimicking clinical trials with synthetic acute myeloid leukemia patients using generative artificial intelligence. *NPJ digital medicine* **2024**, *7*, 76.

75. CC, A.; J, F.; F, M.; M, M.; J, M.; MJ, A.; T, R.; G, M.; M, M. Unlocking the Potential of AI in EUS and ERCP: A Narrative Review for Pancreaticobiliary Disease. *Cancers* **2025**, *17*. https://doi.org/10.3390/cancers1707113 21132.
76. GJ, G.; de Wit NJ.; M, T. Generative artificial intelligence for general practice; new potential ahead, but are we ready? *The European journal of general practice* **2025**, *31*, 2511645. https://doi.org/10.1080/13814788.2025. 25116452511645.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.