

Viewpoint

Multimodal Integration in Health Care: Development With Applications in Disease Management

Yan Hao^{1*}; Chao Cheng^{1*}; Juanjuan Li^{1*}; Hongwen Li²; Xingsi Di³; Xiaoxia Zeng¹; Shoumei Jin⁴; Xiaodong Han⁵; Chongsong Liu¹; Qianqian Wang¹; Bingying Luo⁶; Xianhai Zeng¹; Ke Li¹

¹Department of Otolaryngology, Shenzhen Longgang Otolaryngology Hospital & Shenzhen Otolaryngology Research Institute, Shenzhen, Guangdong, China

²Department of Dentistry, Shenzhen Longgang Otolaryngology Hospital & Shenzhen Otolaryngology Research Institute, Shenzhen, Guangdong, China

³School of Law, Guangzhou University, Guangzhou, Guangdong, China

⁴Department of Ophthalmology, Shenzhen Longgang Otolaryngology Hospital & Shenzhen Otolaryngology Research Institute, Shenzhen, Guangdong, China

⁵Department of Medical Imaging, Shenzhen Longgang Otolaryngology Hospital & Shenzhen Otolaryngology Research Institute, Shenzhen, Guangdong, China

⁶Department of Immunology, Tianjin Medical University, Tianjin, China

*these authors contributed equally

Corresponding Author:

Ke Li

Department of Otolaryngology

Shenzhen Longgang Otolaryngology Hospital & Shenzhen Otolaryngology Research Institute

186 Huangge Road, Longcheng Subdistrict, Longgang District

Shenzhen, Guangdong 518172

China

Phone: 86 (755)28989999

Email: jylike@163.com

Abstract

Multimodal data integration has emerged as a transformative approach in the health care sector, systematically combining complementary biological and clinical data sources such as genomics, medical imaging, electronic health records, and wearable device outputs. This approach provides a multidimensional perspective of patient health that enhances the diagnosis, treatment, and management of various medical conditions. This viewpoint presents an overview of the current state of multimodal integration in health care, spanning clinical applications, current challenges, and future directions. We focus primarily on its applications across different disease domains, particularly in oncology and ophthalmology. Other diseases are briefly discussed due to the few available literature. In oncology, the integration of multimodal data enables more precise tumor characterization and personalized treatment plans. Multimodal fusion demonstrates accurate prediction of anti-human epidermal growth factor receptor 2 therapy response (area under the curve=0.91). In ophthalmology, multimodal integration through the combination of genetic and imaging data facilitates the early diagnosis of retinal diseases. However, substantial challenges remain regarding data standardization, model deployment, and model interpretability. We also highlight the future directions of multimodal integration, including its expanded disease applications, such as neurological and otolaryngological diseases, and the trend toward large-scale multimodal models, which enhance accuracy. Overall, the innovative potential of multimodal integration is expected to further revolutionize the health care industry, providing more comprehensive and personalized solutions for disease management.

J Med Internet Res 2025;27:e76557; doi: [10.2196/76557](https://doi.org/10.2196/76557)

Keywords: multimodal integration; healthcare; personalized medicine; artificial intelligence; digital health

Introduction

In the realm of computer science, the concept of multimodal data refers to the integration and analysis of information from multiple sources or modalities. These modalities can include text, images, audio, video, and sensor data, among others [1]. The primary objective of multimodal data integration is to leverage the complementary strengths of different data types to gain a more comprehensive understanding of a given problem or phenomenon. By combining diverse data sources, multimodal approaches can enhance the accuracy, robustness, and depth of analysis [2,3].

In the context of health care, the application of multimodal data integration becomes even more critical due to the diversity of medical information. The health care sector generates vast amounts of data from a wide array of sources, including medical imaging (such as magnetic resonance imaging [MRI], computed tomography [CT] scans, and x-rays), laboratory test results, electronic health records (EHRs), wearable devices, and environmental sensors [4]. Medical imaging modalities provide detailed anatomical and functional views of the body. EHRs contain a wealth of clinical information, including patient history, diagnoses, treatments, and outcomes, which are essential for longitudinal health monitoring. Wearable devices continuously monitor physiological parameters, such as heart rate, blood pressure, and physical activity, providing real-time data on a patient's health status. Each of these data types provides unique and valuable insights into patient health, but when considered in isolation, they may offer an incomplete or fragmented view. The integration of these diverse data sources enables a more nuanced and comprehensive understanding of patient health.

However, the integration and analysis of multimodal data in health care present significant difficulties. The sheer volume and heterogeneity of the data require sophisticated methodologies capable of handling large, complex datasets. This is where artificial intelligence (AI) and machine learning come into play. The development of multimodal AI is a rapidly evolving field. This approach has already shown promise in various areas of health care [5-7]. Through AI-driven integration of multimodal data, health care providers can achieve a more comprehensive understanding of patient conditions, leading to more accurate diagnoses, personalized treatments, and improved patient outcomes [8].

The future of multimodal integration in health care is promising, with ongoing research and technological

advancements poised to further enhance its capabilities and applications. Emerging technologies, such as advanced imaging modalities, next-generation sequencing, and novel wearable devices, are expected to provide even richer datasets for integration [9]. In addition, the development of more sophisticated AI algorithms and data fusion techniques will enhance the ability to analyze and interpret complex multimodal data.

Despite the vast potential of multimodal integration in health care, several challenges remain to be addressed. First, data standardization and privacy protection require robust solutions while ensuring regulatory compliance. Second, model training and deployment face computational bottlenecks when processing large-scale and biased multimodal datasets. Third, model interpretability must be enhanced to provide clinically meaningful explanations that gain physician trust. Overcoming these barriers is critical for realizing the full clinical potential of multimodal health care systems.

The purpose of this viewpoint is to provide an overview of the current state of multimodal integration in health care, summarize its applications across key disease domains, and discuss the challenges and future directions in this rapidly evolving field. By examining the development and applications of multimodal integration across different disease domains, this viewpoint aims to offer insights into how this approach can further revolutionize the health care industry by providing more comprehensive and personalized solutions for disease management. The content of this study was informed by a systematic search of relevant studies (Multimedia Appendix 1).

Applications

Overview

This section focuses on 2 clinical domains that have seen particularly robust development of multimodal AI applications—oncology and ophthalmology. These specialties were selected due to their substantial body of published research and complex diagnostic requirements benefiting from multimodal data. As summarized in Table 1, we provide a summary of current multimodal developments in these fields.

Table 1. Multimodal artificial intelligence applications across specialties.

Disease and application directions	Specific examples
Oncology	
Enhanced tumor characterization	Tumor subtype and tumor microenvironment
Personalized treatment planning	Personalized radiotherapy and immunotherapy
Early detection and diagnosis	Early cancer detection
Predicting disease prognosis	Overall survival and progression-free survival
Ophthalmology	

Disease and application directions	Specific examples
Early diagnosis and risk stratification	Glaucoma and age-related macular degeneration
Ophthalmology imaging as a noninvasive predictive tool for circulatory system disease	Cardiovascular disease

Application of Multimodal Data in Oncology

Overview

The integration of multimodal data in cancer care represents one of the most promising advancements in modern oncology. For example, advancements in quantitative multimodal imaging technologies involve the combination of multiple quantitative functional measurements, thereby providing a more comprehensive characterization of tumor phenotypes [10]. In addition, integrated genomic analysis methods can reveal dysregulation in biological functions and molecular pathways, offering new opportunities for personalized treatment and monitoring [11]. By combining diverse data sources, health care providers can achieve a more comprehensive understanding of cancer biology, leading to more accurate predictions of patient outcomes. This section explores the various applications of multimodal data in cancer care, highlighting specific case studies and the transformative impact of this approach.

Enhanced Tumor Characterization

One of the primary objectives of integrating multimodal data in cancer care is to achieve enhanced tumor characterization. Tumor characterization involves understanding the genetic, molecular, and phenotypic features of a tumor [12-14], which is essential for elucidating the nature and properties of the malignancy.

A key aspect of this process is the differentiation of tumor subtypes. Tumor subtypes refer to the classification of tumors into distinct categories. Differentiating tumor subtypes is essential because it allows for more precise diagnosis, prognosis, and the development of tailored treatment strategies, specific to the characteristics of each subtype [15]. Previous cancer subtypes were often classified based on gene expression profiles, such as the PAM50 method [16,17]. However, patients within the same group may still experience different outcomes [18], indicating the need for more accurate subtype classification methods. Pathological images and omics data are commonly used for accurate tumor classification through multimodal integration. The features derived from the fusion of image modality data with genomic and other omics data can predict breast cancer subtypes [19]. Typically, dedicated feature extractors are used for each modality. A trained convolutional neural network model captures deep features from pathological images, while a trained deep neural network model extracts features from genomic and other omics data. These multimodal features are then integrated through a fusion model to achieve an accurate prediction of breast cancer molecular subtypes. This integrative approach can also be extended to other tumor types and even pan-cancer studies to support the prediction

of cancer subtypes and severity [20-22]. A large-scale study integrated transcriptome, exome, and pathology data from over 200,000 tumors to develop a multilineage cancer subtype classifier [18].

The tumor microenvironment (TME) plays a crucial role in tumor initiation, progression, metastasis, and resistance to therapy [23,24]. In recent years, advancements in new technologies such as single-cell and spatial technologies [25] have provided fine-grained resolution of TME, significantly enhancing our understanding of cellular interactions at both single-cell and spatial dimensions [26,27]. Besides, the use of multimodal nanosensors can achieve real-time monitoring within the TME [28]. Using multimodal features extracted from single-cell and spatial transcriptomics reveals immunotherapy-relevant non-squamous cell carcinoma (non-small cell lung cancer [NSCLC]) TME heterogeneity [29]. The combination of the 2 modalities and multiplexed ion beam imaging identifies distinct tumor subgroups and a cancer-specific tumor-specific keratinocyte [30]. Spatial multiomics delineate core and margin compartments in oral squamous cell carcinoma, with metabolically active margins demonstrating elevated adenosine triphosphate production to fuel invasion [31]. In cross-modal applications, gene expression can be predicted from histopathological images of breast cancer tissue with a resolution of 100 μm [32]. Conversely, spatial transcriptomic features can better characterize breast cancer tissue sections, revealing hidden histological features [33]. By extracting interpretable features from pathological slides, it is also possible to predict different molecular phenotypes [34]. These methods provide a comprehensive, quantitative, and interpretable window into the composition and spatial structure of the TME.

Personalized Treatment Planning

Another critical objective of multimodal data integration in cancer care is personalized treatment planning. Personalized treatment involves tailoring medical interventions to the individual characteristics of each patient, taking into account their tumor biology and overall health status. By integrating data from multiple sources, health care providers can develop more precise and personalized treatment plans that improve patient outcomes.

In terms of radiation therapy, using multimodal scanning techniques and mathematical models, it is possible to design personalized radiotherapy plans for glioblastoma patients. By integrating high-resolution MRI scans and metabolic profiles, this approach enables more accurate inference of tumor cell density, thereby optimizing radiotherapy regimens and reducing damage to healthy tissue [35]. The integration of biological information-driven multimodal imaging techniques allows physicians to better understand the spatial

and temporal heterogeneity of tumors to develop personalized radiotherapy regimens [36].

In the trend of precision medicine, another therapeutic approach is immunotherapy. Immune checkpoint blockade can unleash immune cells to reinvigorate antitumor immunity [37]. Multiple phase III clinical trials have demonstrated that the anti-programmed cell death protein 1 antibody nivolumab significantly improves overall survival with a favorable safety profile in patients with NSCLC [38]. Although single-modality biomarkers can predict responses to immune checkpoint blockade, their predictive power is not always satisfactory. Activating an antitumor immune response through immunotherapy involves a series of complex events that require the interaction of multiple cell types [39]. Therefore, achieving precision immunotherapy necessitates integrating multiple data modalities and adopting a holistic approach to analyze the human TME. Translating these multimodal factors into clinically usable predictive markers facilitates the selection of optimal immunotherapy. Combining the informational content present in routine diagnostic data, including annotated CT scans, digitized immunohistochemistry slides, and common genomic alterations in NSCLC, can improve the prediction of responses to programmed cell death protein 1 or programmed cell death-ligand 1 blockade [40]. Multimodal model by Chen et al [41] can predict the response to anti-human epidermal growth factor receptor 2 combined immunotherapy using multimodal radiology, pathology, and clinical information, achieving an area under the curve (AUC) of 0.91. Furthermore, the application of multimodal approaches in targeted cancer therapy has demonstrated significant potential. Integrating radiomic phenotypes with liquid biopsy data can enhance the predictive accuracy for the efficacy of epidermal growth factor receptor inhibitors [42].

Early Detection and Diagnosis

Early detection and diagnosis of cancer are crucial for improving patient outcomes, as early-stage cancers are often more treatable and have better prognoses [43]. Multimodal data integration plays a vital role in enhancing the accuracy and timeliness of cancer detection and diagnosis.

Liquid biopsy is a noninvasive technique that involves the collection of nonsolid samples, providing possibilities for early cancer detection and longitudinal tracking [44]. This technology includes circulating tumor cells shed from primary and metastatic tumors, as well as circulating tumor DNA (ctDNA) [45]. ctDNA can detect trace amounts of tumor DNA even before the tumor manifests obvious symptoms or becomes visible through imaging. Numerous studies and articles have used ctDNA in combination with various other modalities for early cancer prediction, including lung cancer [46], breast cancer [47], and colorectal cancer [48]. Cell-free DNA is a substance that is consistently present in plasma and has been receiving increasing attention. Combining cell-free DNA with other modalities can be used for highly specific early detection across multiple cancer types [49-51]. AutoCancer uses a transformer model to integrate multiple modalities, including liquid biopsy, mutation, and clinical data, achieving accurate early cancer

detection in both lung cancer and pan-cancer analyses [52]. Multimodal models that integrate genomic features and clinical data have also demonstrated excellent performance in the early detection of colorectal cancer, with an AUC of 0.98 in the validation set and a sensitivity and specificity of more than 90% [49].

Predicting Disease Prognosis

Prognosis involves assessing the risk of future outcomes based on an individual's clinical and nonclinical characteristics. These outcomes are typically specific events, such as death or complications, but they can also be quantitative measures, such as disease progression, changes in pain levels, or quality of life [53]. Predicting disease prognosis is a critical aspect of cancer care, as it allows for timely interventions and improved long-term outcomes. Multimodal data integration enhances the ability to predict disease prognosis.

Prognosis in tumor research can be divided into 2 key areas: recurrence and survival. In the context of recurrence, a retrospective analysis and multicenter validation study involving over 2000 patients demonstrated that a multimodal recurrence score, which integrated clinical, genomic, and histopathological data, accurately predicted postoperative local recurrence of renal cell carcinoma [54]. Combining the emerging tool of habitat imaging with traditional gene expression and clinical data enables noninvasive stratification of patients with NSCLC, enhancing the prediction of recurrence risk [55]. In another study, algorithms were developed based on structured clinical and administrative data to detect recurrence in lung and colorectal cancer patients. By using EHRs and tumor registry data, these algorithms successfully improved the accuracy of recurrence detection [56].

Regarding survival, an increasing number of studies have adopted multimodal approaches to predict patient survival [57-61]. By integrating data from various sources, these studies have achieved accurate survival predictions across multiple tumor types, including overall survival, 5-year survival rates, and progression-free survival.

Application of Multimodal Data in Ophthalmology

Overview

Ophthalmology, the medical specialty focused on the diagnosis and treatment of eye disorders, has experienced significant advancements through the integration of multimodal data. Advanced imaging techniques are central to ophthalmology, providing detailed visualizations of the retina, optic nerve, and other ocular structures [62]. Optical coherence tomography (OCT) is a widely used imaging modality that offers high-resolution cross-sectional images of the retina, enabling the detection of structural abnormalities and disease progression. Fundus photography and fluorescein angiography provide additional insights into the retinal vasculature and blood flow, which are critical for diagnosing and managing conditions like diabetic retinopathy and retinal vein occlusion. These imaging techniques, when integrated,

offer a comprehensive view of both the structural and genetic factors contributing to ocular diseases. The fusion of these data types enables early diagnosis, personalized treatment plans, and continuous monitoring of disease progression and response to therapy, particularly in conditions like age-related macular degeneration (AMD), diabetic retinopathy, and glaucoma [63].

Early Diagnosis and Risk Stratification

The integration of these diverse data types in ophthalmology achieves several important objectives. Early diagnosis and risk stratification are critical for managing ocular diseases, and the combination of genetic, imaging, and clinical data enables the identification of early signs of eye conditions and stratification of patients based on their risk profiles.

Color fundus photography and OCT are 2 of the most cost-effective tools for glaucoma screening. Mehta et al [64] developed a high-performance multimodal glaucoma detection system by integrating OCT volumes, fundus photographs, and clinical data. Their approach combined features extracted from individual modalities, followed by gradient boosting decision trees for final multimodal construction. The model was rigorously developed and validated on a cohort of 96,020 UK Biobank participants, demonstrating both excellent discriminative performance (AUC=0.97). Importantly, the architecture maintained clinical interpretability through comprehensive feature importance analysis [64]. Other multimodal models for glaucoma and its grading detection, based on modalities, such as OCT and fundus images, have also achieved AUC exceeding 0.90 [65-67]. By using a dual-stream convolutional neural network model to extract features from OCT and color fundus photographs, AMD can be classified into 3 categories—normal fundus, dry AMD, and wet AMD [68]. Another study enrolled 75 participants from optometry clinics in Auckland and Milford Eye Clinic, New Zealand. By stratifying subjects into young healthy controls, older adult healthy controls, and moderate dry AMD groups, the multimodal diagnostic system achieved 96% classification accuracy [69]. In addition, the use of multimodal data can also identify polypoidal choroidal vasculopathy [70], dry eye disease [71], and diabetic retinopathy [72-75]. There is also comprehensive work demonstrating that multimodal deep learning (DL) models, which use combined color fundus photography and OCT image sequences as input, can be used to simultaneously detect multiple common retinal diseases [76,77].

Ophthalmology Imaging as a Noninvasive Predictive Tool for Circulatory System Disease

Currently, the diagnosis and treatment of circulatory system disease primarily rely on imaging examinations such as MRI, coronary CT angiography, and coronary angiography. These examinations are not only expensive and time-consuming but also partially invasive and require a high level of professional expertise from the operators. Consequently, early screening and long-term follow-up examinations are challenging to implement in regions with limited medical resources. To better achieve early warning and assessment

of circulatory system disease, there is a continuous need to develop new diagnostic tools that are noninvasive, convenient, and efficient.

The microcirculation of the retina is part of the body's microcirculation system and shares similar embryological origins and pathophysiological characteristics with the cardiovascular system [78]. Numerous studies have identified retinal imaging biomarkers associated with early cardiovascular diseases (CVDs) lesions and prognosis, demonstrating the significant value of retinal imaging in CVD screening and prognostic evaluation [79,80].

Al-Absi et al [81] used a multimodal approach integrating retinal images and dual-energy x-ray absorptiometry data to diagnose CVD in a Qatari cohort. The multimodal model achieved 78.3% accuracy, outperforming unimodal models [81]. Notably, their model is interpretable, using Gradient-weighted Class Activation Mapping (Grad-CAM) to highlight the areas of interest in retinal images that most influenced the decisions of the proposed DL model. A study using clinical information and fundus photographs from the UK Biobank demonstrated a significant association between the incidence of CVD in high-risk patients and multimodal predicted risk (hazard ratio 6.28, 95% CI 4.72-8.34), and visualized feature importance [82].

Challenges in Multimodal Health Care

While the integration of multimodal data in health care holds great promise, it also presents several significant challenges that need to be addressed.

Data Standardization and Privacy

One of the primary challenges in multimodal health care is integrating diverse medical data sources with varying formats, resolutions, and quality levels [83]. Inconsistent data collection practices, missing entries, and recording errors can compromise model reliability, necessitating robust standardization protocols [84]. Effective multimodal integration requires comprehensive data cleaning, validation, and preprocessing to create cohesive, high-quality datasets that support accurate predictive analytics. The growing availability of novel health care data sources presents both opportunities for personalized medicine and challenges for systematic integration.

The use of multimodal data in health care raises significant concerns about data privacy and security. Medical data is highly sensitive, and ensuring its protection is paramount. Regulatory frameworks, such as the Health Insurance Portability and Accountability Act in the United States and the General Data Protection Regulation in the European Union, are essential for protecting patient privacy and ensuring data security. But the concept of health information privacy continues to evolve over time. As new technologies and data sources emerge, it is essential to update and adapt these legal frameworks to reflect new realities [85]. The use of multimodal data raises significant privacy concerns [86].

Implementing robust data encryption, secure data storage, and strict access controls are essential measures to protect patient information [87]. Comprehensive data governance frameworks must establish clear guidelines for responsible and transparent multimodal data usage, while carefully balancing potential risks and benefits for participants, researchers, and society at large [88]. Effective implementation requires developing robust data sharing agreements, establishing independent oversight committees, and maintaining ongoing engagement with research participants and other stakeholders [89]. In addition, developing secure data sharing protocols and anonymization techniques can help mitigate risks while enabling the effective use of multimodal data for research and clinical applications [90]. Ensuring data privacy and security is fundamental to maintaining patient trust and the ethical use of medical data.

The initial phase of multimodal health care requires systematic collection of standardized data following heterogeneity resolution, coupled with privacy protection through secure protocols. Integrating rigorous data processing with ethically compliant governance frameworks enables usage of diverse datasets for precision medicine while safeguarding sensitive information. This equilibrium is critical for advancing research ethically and maintaining public trust in medical AI applications.

Model Training and Deployment

Multimodal models demand substantial computational resources for both training and inference. The complexity of these models often results in extended training times and significant costs, which can be prohibitive for many health care institutions [91,92]. Training these models requires high-performance computing environments equipped with powerful Graphics Processing Units or Tensor Processing Units, which are not always accessible to all institutions. Furthermore, the inference phase, where the trained model is applied to new data, can also be resource-intensive, particularly when dealing with large-scale datasets or real-time applications [93]. This computational burden can limit the scalability and practical deployment of multimodal models in clinical settings.

Beyond computational constraints, biases in training data pose a significant challenge to multimodal fusion. Biases may arise from uneven data distribution, inconsistent annotation quality, or systemic disparities in data collection. AI-driven decisions are fundamentally shaped by their initial training data. If the underlying datasets contain biases or inequities, the resulting algorithms risk perpetuating prejudice, incomplete representations, or discriminatory outcomes—potentially amplifying systemic inequalities [94]. To counteract these biases, strategies such as bias-aware sampling and fairness constraints during model optimization can be implemented. While some AI developers claim their algorithmic systems can mitigate biases, critics maintain that algorithms alone cannot eradicate discrimination, as they may inadvertently perpetuate existing bias in training data [95]. This tension highlights the need for complementary

strategies (ie, rigorous dataset curation to ensure diversity and continuous monitoring for disparate impacts) [96].

Training and running multimodal models demand expensive hardware, limiting clinical adoption. Meanwhile, biased training data can perpetuate health care disparities. While optimization techniques and bias mitigation strategies help, robust data curation and ongoing monitoring of potentially biased data remain essential for practical, equitable deployment.

Model Interpretability

While multimodal models can achieve high accuracy, their complexity often makes them difficult to interpret. This lack of interpretability poses a significant barrier to their adoption in clinical practice, as clinicians and patients need to understand the rationale behind model predictions to trust and effectively use these tools [97]. Enhancing the interpretability and transparency of multimodal models is therefore crucial [98]. Techniques, such as explainable artificial intelligence (XAI), can play a pivotal role in this regard [99]. XAI methods aim to make the decision-making processes of AI models more understandable to humans by providing explanations that are both accurate and comprehensible. Classical XAI approaches include attention mechanisms and Grad-CAM. Attention scores highlight relevant regions through forward propagation, while Grad-CAM reveals feature significance by capturing gradient changes during backpropagation [100].

Attention mechanisms were originally developed to help neural networks focus on the most relevant parts of input data when making predictions. The core principle involves calculating attention weights—numerical scores that determine how much each input element (eg, words in text or regions in an image) should influence the model's output [101]. MedFuseNet [102] uses an image attention mechanism to dynamically focus on the most clinically relevant regions of medical images corresponding to the input textual queries. Visualization of the attention matrices reveals that the model consistently attends to anatomically discriminative regions of target organs, demonstrating its capability to identify pathologically significant features. StereoMM [103] enables quantitative analysis of cross-attention matrices to determine the relative contribution weights of different modalities during fusion, thereby offering interpretable insights into the prioritization of modalities by the model in its decision-making process. Nevertheless, attention weights primarily reflect statistical correlations rather than causal relationships. The fact that a feature receives high attention does not necessarily imply it was determinative for the model's prediction. Compounding this issue, empirical studies have demonstrated that substantially different attention weight distributions can yield identical model outputs [104]. These limitations raise questions about the validity of using attention mechanisms as reliable tools for explaining neural network behavior, making an ongoing subject of debate in the machine learning community [105].

Grad-CAM generates explanations by computing gradients from the final convolutional layer, highlighting prediction-relevant regions [106]. This interpretability method helps detect invalid decision patterns. For instance, if highest activations appear on imaging artifacts rather than anatomical structures, it exposes critical model flaws. In a clinical study using brain MRI for classification of multiple sclerosis subtypes, Grad-CAM-generated heatmaps consistently and distinctly highlighted brain regions critical for differentiating between subtypes, thereby demonstrating the validity and explanatory power. Furthermore, Grad-CAM analysis identified previously unrecognized neuroanatomical loci, offering novel insights into disease progression mechanisms and potentially revealing new imaging biomarkers or therapeutic targets [107]. It should be noted that Grad-CAM offers qualitative visualization of model decisions, not quantitative validation. Its clinical relevance must be determined through physician assessment of the identified features [108].

Multimodal AI models face a key challenge—balancing high accuracy with clinical interpretability. Current XAI methods offer partial solutions, but with important limitations. Both methods produce explanations that require clinical validation, and physician expertise remains essential to assess biological plausibility. These limitations highlight the need for XAI approaches that provide both technical transparency and clinically meaningful explanations to enable trustworthy AI adoption in health care.

The Development Direction of Multimodal Technology: Expanding Disease Applications

The development of multimodal technology encompasses broader applications across various diseases and the advancement of large-scale models. With technological progress, multimodal approaches are no longer limited to the diagnosis and prognosis of cancer and ophthalmic diseases but are expanding into CVD, neurological disorders, metabolic diseases, otolaryngology, and more.

In the field of CVD, multimodal technology can combine data from cardiac MRI, coronary CT, echocardiography, and biomarkers to provide a more comprehensive assessment of heart health [87]. For example, integrating these data can more accurately predict the risk of ischemic heart disease [109,110], coronary artery disease [111], assess cardiac function [112], and detect disease subgroups plans [113]. In addition, multimodal technology can be used to monitor the treatment effects and disease progression in heart disease patients, allowing timely adjustments to treatment strategies and improving patient survival rates and quality of life [114].

In the realm of neurological disorders, multimodal technology also holds significant promise. A proposed model demonstrates robust multimodal integration capabilities, effectively combining both imaging and nonimaging clinical data to achieve accurate differential diagnosis of Alzheimer disease, with discriminative performance exceeding AUC values of 0.9 across multiple diagnostic tasks [115]. By

combining brain MRI, functional MRI, electroencephalography, and genomic data, researchers can gain a more comprehensive understanding of the pathophysiology of diseases such as Alzheimer [116], Parkinson [117], and multiple sclerosis [118]. Integrating these data can aid in the early diagnosis of these diseases and assess disease severity.

In the field of metabolic diseases, multimodal technology also has important applications. Integrating clinical documentation with structured laboratory data significantly improves the predictive performance of unimodal machine learning models for early-stage type 2 diabetes mellitus detection. The model achieved an AUC greater than 0.70 for new-onset type 2 diabetes mellitus prediction [119]. By integrating metabolomics, genomics, imaging, and clinical data, researchers can gain a more comprehensive understanding of the pathophysiology of diseases, such as obesity [120] and fatty liver disease [121]. Integrating these data can aid in the early diagnosis of these diseases and assess disease status.

In the field of otolaryngology, the automatic classification of parotid gland tumors based on multimodal MRI sequences shows promise for improving diagnostic decision-making in clinical settings [122]. The integration of CT and MRI enables precise tumor segmentation of oropharyngeal squamous cell carcinoma, resulting in higher dice similarity coefficients and lower Hausdorff distances [123]. Combining otoscopic images and wideband tympanometry enables the automatic detection of otitis media [124]. Institutions have recognized the importance of collecting multimodal data for interdisciplinary audiology research and have developed a multimodal database that can be used for algorithm development [125].

The Trend Toward Large Language Models

Large language models (LLMs) are foundational pretrained AI systems capable of processing and generating human-like text [126]. Their key advantage lies in capturing complex semantic relationships within language data. Building upon LLMs, large multimodal models extend these capabilities to integrate and analyze diverse data types (text, images, genomic data, etc), achieving significant advancements and breakthroughs, gradually forming the rudiments of artificial general intelligence [127]. The trend toward LLM in multimodal technology enhances the accuracy and robustness of disease prediction and diagnosis by capturing complex relationships between different data types [128,129].

For example, transformer models, which have achieved remarkable success in natural language processing and computer vision, are now being applied to the integration and analysis of multimodal data [130]. The transformer-based unified multimodal diagnostic transformer model is capable of directly generating diagnostic results for lung diseases from multimodal input data [131].

Furthermore, LLMs have stronger generalization capabilities, allowing them to be applied across various

diseases and populations. This general-purpose approach not only enhances diagnostic accuracy but also reduces the cost and complexity of training and deploying multiple specialized models. For instance, a single large multimodal model could be used for the diagnosis and prognosis of cancer, aging and age-related diseases [132], CVDs, neurological disorders, and metabolic diseases, streamlining the process and improving efficiency.

Another important aspect of LLMs is their interpretability, primarily achieved through the use of attention weights. Although DL models are often considered “black boxes,” recent advancements have focused on improving model transparency. Attention mechanisms enhance interpretability

by identifying and emphasizing the most critical features in the input data, allowing attention to be visualized as regions of information that contribute to decision-making [133,134]. By visualizing the distribution of attention weights, one can extract the content with high attention weights, which often have a greater impact on the final outcome prediction [135].

In summary, the trend toward LLMs in multimodal development is poised to bring significant innovations and breakthroughs to the medical field. By leveraging the power of large-scale, multimodal datasets and advanced neural network architectures, researchers can achieve more accurate and comprehensive disease predictions and diagnoses.

Acknowledgments

This work was supported by Shenzhen Science and Technology Plan Projects (JCYJ20220530154200002 and JCYJ20230807091701004), Shenzhen Key Medical Discipline Construction Fund (SZXK039), and Longgang District Medical and Health Technology Attack Project (LGKCYLWS2023027).

Authors' Contributions

YH, CC, JL, XH, BL, and SJ finished the writing-original draft. HL and Xianhai Z were involved in investigation and validation. CL, XD, and QW did conceptualization and editing. CC, Xianhai Z, and KL performed supervision and funding acquisition.

Xianhai Z is the co-corresponding author of this paper and can be reached at: Department of Otolaryngology, Shenzhen Longgang Otolaryngology Hospital & Shenzhen Otolaryngology Research Institute; zxhklwx@163.com

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional material.

[DOCX File (Microsoft Word File), 18 KB-Multimedia Appendix 1]

References

1. Baltrusaitis T, Ahuja C, Morency LP. Multimodal machine learning: a survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell*. Feb 2019;41(2):423-443. [doi: [10.1109/TPAMI.2018.2798607](https://doi.org/10.1109/TPAMI.2018.2798607)] [Medline: [29994351](https://pubmed.ncbi.nlm.nih.gov/29994351/)]
2. Xu X, Li J, Zhu Z, et al. A comprehensive review on synergy of multi-modal data and AI technologies in medical diagnosis. *Bioengineering (Basel)*. Feb 25, 2024;11(3):219. [doi: [10.3390/bioengineering11030219](https://doi.org/10.3390/bioengineering11030219)] [Medline: [38534493](https://pubmed.ncbi.nlm.nih.gov/38534493/)]
3. Atrey PK, Hossain MA, El Saddik A, Kankanhalli MS. Multimodal fusion for multimedia analysis: a survey. *Multimedia Systems*. Nov 2010;16(6):345-379. [doi: [10.1007/s00530-010-0182-0](https://doi.org/10.1007/s00530-010-0182-0)]
4. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. *J Big Data*. Dec 2019;6(1):54. [doi: [10.1186/s40537-019-0217-0](https://doi.org/10.1186/s40537-019-0217-0)]
5. Zhao AP, Li S, Cao Z, et al. AI for science: predicting infectious diseases. *Journal of Safety Science and Resilience*. Jun 2024;5(2):130-146. [doi: [10.1016/j.jnlssr.2024.02.002](https://doi.org/10.1016/j.jnlssr.2024.02.002)]
6. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. *Bioengineering (Basel)*. Dec 18, 2023;10(12):1435. [doi: [10.3390/bioengineering10121435](https://doi.org/10.3390/bioengineering10121435)] [Medline: [38136026](https://pubmed.ncbi.nlm.nih.gov/38136026/)]
7. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med*. Sep 2022;28(9):1773-1784. [doi: [10.1038/s41591-022-01981-2](https://doi.org/10.1038/s41591-022-01981-2)] [Medline: [36109635](https://pubmed.ncbi.nlm.nih.gov/36109635/)]
8. Moghadam MP, Moghadam ZA, Qazani MRC, Pławiak P, Alizadehsani R. Impact of artificial intelligence in nursing for geriatric clinical care for chronic diseases: a systematic literature review. *IEEE Access*. 2024;12:122557-122587. [doi: [10.1109/ACCESS.2024.3450970](https://doi.org/10.1109/ACCESS.2024.3450970)]
9. Shaik T, Tao X, Li L, Xie H, Velásquez JD. A survey of multimodal information fusion for smart healthcare: mapping the journey from data to wisdom. *Information Fusion*. Feb 2024;102:102040. [doi: [10.1016/j.inffus.2023.102040](https://doi.org/10.1016/j.inffus.2023.102040)]
10. Yankeelov TE, Abramson RG, Quarles CC. Quantitative multimodality imaging in cancer research and therapy. *Nat Rev Clin Oncol*. Nov 2014;11(11):670-680. [doi: [10.1038/nrclinonc.2014.134](https://doi.org/10.1038/nrclinonc.2014.134)] [Medline: [25113842](https://pubmed.ncbi.nlm.nih.gov/25113842/)]

11. Kristensen VN, Lingjærde OC, Russnes HG, Volla HMK, Frigessi A, Børresen-Dale AL. Principles and methods of integrative genomic analyses in cancer. *Nat Rev Cancer*. May 2014;14(5):299-313. [doi: [10.1038/nrc3721](https://doi.org/10.1038/nrc3721)] [Medline: [24759209](https://pubmed.ncbi.nlm.nih.gov/24759209/)]
12. Liu Z, Zhang S. Tumor characterization and stratification by integrated molecular profiles reveals essential pan-cancer features. *BMC Genomics*. Jul 7, 2015;16(1):503. [doi: [10.1186/s12864-015-1687-x](https://doi.org/10.1186/s12864-015-1687-x)] [Medline: [26148869](https://pubmed.ncbi.nlm.nih.gov/26148869/)]
13. Jena B, Saxena S, Nayak GK, et al. Brain tumor characterization using radiogenomics in artificial intelligence framework. *Cancers (Basel)*. Aug 22, 2022;14(16):4052. [doi: [10.3390/cancers14164052](https://doi.org/10.3390/cancers14164052)] [Medline: [36011048](https://pubmed.ncbi.nlm.nih.gov/36011048/)]
14. Hoffmann E, Masthoff M, Kunz WG, et al. Multiparametric MRI for characterization of the tumour microenvironment. *Nat Rev Clin Oncol*. Jun 2024;21(6):428-448. [doi: [10.1038/s41571-024-00891-1](https://doi.org/10.1038/s41571-024-00891-1)] [Medline: [38641651](https://pubmed.ncbi.nlm.nih.gov/38641651/)]
15. Yeo SK, Guan JL. Breast cancer: multiple subtypes within a tumor? *Trends Cancer*. Nov 2017;3(11):753-760. [doi: [10.1016/j.trecan.2017.09.001](https://doi.org/10.1016/j.trecan.2017.09.001)] [Medline: [29120751](https://pubmed.ncbi.nlm.nih.gov/29120751/)]
16. Pu M, Messer K, Davies SR, et al. Research-based PAM50 signature and long-term breast cancer survival. *Breast Cancer Res Treat*. Jan 2020;179(1):197-206. [doi: [10.1007/s10549-019-05446-y](https://doi.org/10.1007/s10549-019-05446-y)]
17. Parker JS, Mullins M, Cheang MCU, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. Mar 10, 2009;27(8):1160-1167. [doi: [10.1200/JCO.2008.18.1370](https://doi.org/10.1200/JCO.2008.18.1370)] [Medline: [19204204](https://pubmed.ncbi.nlm.nih.gov/19204204/)]
18. Shergalis A, Bankhead A III, Luesakul U, Muangsin N, Neamati N. Current challenges and opportunities in treating glioblastoma. *Pharmacol Rev*. Jul 2018;70(3):412-445. [doi: [10.1124/pr.117.014944](https://doi.org/10.1124/pr.117.014944)] [Medline: [29669750](https://pubmed.ncbi.nlm.nih.gov/29669750/)]
19. Liu T, Huang J, Liao T, Pu R, Liu S, Peng Y. A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data. *IRBM*. Feb 2022;43(1):62-74. [doi: [10.1016/j.irbm.2020.12.002](https://doi.org/10.1016/j.irbm.2020.12.002)]
20. Duroux D, Wohlfart C, Van Steen K, Vladimirova A, King M. Graph-based multi-modality integration for prediction of cancer subtype and severity. *Sci Rep*. Nov 10, 2023;13(1):19653. [doi: [10.1038/s41598-023-46392-6](https://doi.org/10.1038/s41598-023-46392-6)] [Medline: [37949935](https://pubmed.ncbi.nlm.nih.gov/37949935/)]
21. Ding S, Li J, Wang J, Ying S, Shi J. Multimodal co-attention fusion network with online data augmentation for cancer subtype classification. *IEEE Trans Med Imaging*. Nov 2024;43(11):3977-3989. [doi: [10.1109/TMI.2024.3405535](https://doi.org/10.1109/TMI.2024.3405535)] [Medline: [38801690](https://pubmed.ncbi.nlm.nih.gov/38801690/)]
22. Li B, Nabavi S. A multimodal graph neural network framework for cancer molecular subtype classification. *BMC Bioinformatics*. Jan 15, 2024;25(1):27. [doi: [10.1186/s12859-023-05622-4](https://doi.org/10.1186/s12859-023-05622-4)] [Medline: [38225583](https://pubmed.ncbi.nlm.nih.gov/38225583/)]
23. Anderson NM, Simon MC. The tumor microenvironment. *Curr Biol*. Aug 17, 2020;30(16):R921-R925. [doi: [10.1016/j.cub.2020.06.081](https://doi.org/10.1016/j.cub.2020.06.081)] [Medline: [32810447](https://pubmed.ncbi.nlm.nih.gov/32810447/)]
24. Baghban R, Roshangar L, Jahanban-Esfahlan R, et al. Tumor microenvironment complexity and therapeutic implications at a glance. *Cell Commun Signal*. Apr 7, 2020;18(1):59. [doi: [10.1186/s12964-020-0530-4](https://doi.org/10.1186/s12964-020-0530-4)] [Medline: [32264958](https://pubmed.ncbi.nlm.nih.gov/32264958/)]
25. Walsh LA, Quail DF. Decoding the tumor microenvironment with spatial technologies. *Nat Immunol*. Dec 2023;24(12):1982-1993. [doi: [10.1038/s41590-023-01678-9](https://doi.org/10.1038/s41590-023-01678-9)] [Medline: [38012408](https://pubmed.ncbi.nlm.nih.gov/38012408/)]
26. Schürch CM, Bhate SS, Barlow GL, et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell*. Sep 3, 2020;182(5):1341-1359. [doi: [10.1016/j.cell.2020.07.005](https://doi.org/10.1016/j.cell.2020.07.005)] [Medline: [32763154](https://pubmed.ncbi.nlm.nih.gov/32763154/)]
27. Sun C, Wang A, Zhou Y, et al. Spatially resolved multi-omics highlights cell-specific metabolic remodeling and interactions in gastric cancer. *Nat Commun*. May 10, 2023;14(1):37164975. [doi: [10.1038/s41467-023-38360-5](https://doi.org/10.1038/s41467-023-38360-5)]
28. Hao L, Rohani N, Zhao RT, et al. Microenvironment-triggered multimodal precision diagnostics. *Nat Mater*. Oct 2021;20(10):1440-1448. [doi: [10.1038/s41563-021-01042-y](https://doi.org/10.1038/s41563-021-01042-y)] [Medline: [34267368](https://pubmed.ncbi.nlm.nih.gov/34267368/)]
29. Lapuente-Santana Ó, Sturm G, Kant J, et al. Multimodal analysis unveils tumor microenvironment heterogeneity linked to immune activity and evasion. *iScience*. Aug 16, 2024;27(8):110529. [doi: [10.1016/j.isci.2024.110529](https://doi.org/10.1016/j.isci.2024.110529)] [Medline: [39161957](https://pubmed.ncbi.nlm.nih.gov/39161957/)]
30. Ji AL, Rubin AJ, Thrane K, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*. Jul 23, 2020;182(2):497-514. [doi: [10.1016/j.cell.2020.05.039](https://doi.org/10.1016/j.cell.2020.05.039)] [Medline: [32579974](https://pubmed.ncbi.nlm.nih.gov/32579974/)]
31. Arora R, Cao C, Kumar M, et al. Spatial transcriptomics reveals distinct and conserved tumor core and edge architectures that predict survival and targeted therapy response. *Nat Commun*. Aug 18, 2023;14(1):37596273. [doi: [10.1038/s41467-023-40271-4](https://doi.org/10.1038/s41467-023-40271-4)]
32. He B, Bergensträhle L, Stenbeck L, et al. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng*. Aug 2020;4(8):827-834. [doi: [10.1038/s41551-020-0578-x](https://doi.org/10.1038/s41551-020-0578-x)]
33. Monjo T, Koido M, Nagasawa S, Suzuki Y, Kamatani Y. Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. *Sci Rep*. Mar 8, 2022;12(1):35260632. [doi: [10.1038/s41598-022-07685-4](https://doi.org/10.1038/s41598-022-07685-4)]

34. Diao JA, Wang JK, Chui WF, et al. Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nat Commun*. Mar 12, 2021;12(1):1613. [doi: [10.1038/s41467-021-21896-9](https://doi.org/10.1038/s41467-021-21896-9)] [Medline: [33712588](https://pubmed.ncbi.nlm.nih.gov/33712588/)]
35. Lipkova J, Angelikopoulos P, Wu S, et al. Personalized radiotherapy design for glioblastoma: integrating mathematical tumor models, multimodal scans, and Bayesian inference. *IEEE Trans Med Imaging*. Aug 2019;38(8):1875-1884. [doi: [10.1109/TMI.2019.2902044](https://doi.org/10.1109/TMI.2019.2902044)] [Medline: [30835219](https://pubmed.ncbi.nlm.nih.gov/30835219/)]
36. Breen WG, Aryal MP, Cao Y, Kim MM. Integrating multi-modal imaging in radiation treatments for glioblastoma. *Neuro-oncology*. Mar 4, 2024;26(Supplement_1):S17-S25. [doi: [10.1093/neuonc/noad187](https://doi.org/10.1093/neuonc/noad187)]
37. He X, Xu C. Immune checkpoint signaling and cancer immunotherapy. *Cell Res*. Aug 2020;30(8):660-669. [doi: [10.1038/s41422-020-0343-4](https://doi.org/10.1038/s41422-020-0343-4)]
38. Vokes EE, Ready N, Felip E, et al. Nivolumab versus docetaxel in previously treated advanced non-small-cell lung cancer (CheckMate 017 and CheckMate 057): 3-year update and outcomes in patients with liver metastases. *Ann Oncol*. Apr 1, 2018;29(4):959-965. [doi: [10.1093/annonc/mdy041](https://doi.org/10.1093/annonc/mdy041)] [Medline: [29408986](https://pubmed.ncbi.nlm.nih.gov/29408986/)]
39. Roelofsen LM, Kaptein P, Thommen DS. Multimodal predictors for precision immunotherapy. *Immuno-Oncology and Technology*. Jun 2022;14(100071):100071. [doi: [10.1016/j.iotech.2022.100071](https://doi.org/10.1016/j.iotech.2022.100071)]
40. Vanguri RS, Luo J, Aukerman AT, et al. Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. *Nat Cancer*. Oct 2022;3(10):1151-1164. [doi: [10.1038/s43018-022-00416-8](https://doi.org/10.1038/s43018-022-00416-8)] [Medline: [36038778](https://pubmed.ncbi.nlm.nih.gov/36038778/)]
41. Chen Z, Chen Y, Sun Y, et al. Predicting gastric cancer response to anti-HER2 therapy or anti-HER2 combined immunotherapy based on multi-modal data. *Signal Transduct Target Ther*. Aug 26, 2024;9(1):222. [doi: [10.1038/s41392-024-01932-y](https://doi.org/10.1038/s41392-024-01932-y)] [Medline: [39183247](https://pubmed.ncbi.nlm.nih.gov/39183247/)]
42. Yousefi B, LaRiviere MJ, Cohen EA, et al. Combining radiomic phenotypes of non-small cell lung cancer with liquid biopsy data may improve prediction of response to EGFR inhibitors. *Sci Rep*. May 11, 2021;11(1):9984. [doi: [10.1038/s41598-021-88239-y](https://doi.org/10.1038/s41598-021-88239-y)] [Medline: [33976268](https://pubmed.ncbi.nlm.nih.gov/33976268/)]
43. Crosby D, Bhatia S, Brindle KM, et al. Early detection of cancer. *Science*. Mar 18, 2022;375(6586):eaay9040. [doi: [10.1126/science.aay9040](https://doi.org/10.1126/science.aay9040)] [Medline: [35298272](https://pubmed.ncbi.nlm.nih.gov/35298272/)]
44. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol*. Aug 2013;10(8):472-484. [doi: [10.1038/nrclinonc.2013.110](https://doi.org/10.1038/nrclinonc.2013.110)] [Medline: [23836314](https://pubmed.ncbi.nlm.nih.gov/23836314/)]
45. Lone SN, Nisar S, Masoodi T, et al. Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. *Mol Cancer*. Mar 18, 2022;21(1):79. [doi: [10.1186/s12943-022-01543-7](https://doi.org/10.1186/s12943-022-01543-7)] [Medline: [35303879](https://pubmed.ncbi.nlm.nih.gov/35303879/)]
46. Chabon JJ, Hamilton EG, Kurtz DM, et al. Integrating genomic features for non-invasive early lung cancer detection. *Nature New Biol*. Apr 2020;580(7802):245-251. [doi: [10.1038/s41586-020-2140-0](https://doi.org/10.1038/s41586-020-2140-0)] [Medline: [32269342](https://pubmed.ncbi.nlm.nih.gov/32269342/)]
47. Pham TMQ, Phan TH, Jasmine TX, et al. Multimodal analysis of genome-wide methylation, copy number aberrations, and end motif signatures enhances detection of early-stage breast cancer. *Front Oncol*. 2023;13(1127086):1127086. [doi: [10.3389/fonc.2023.1127086](https://doi.org/10.3389/fonc.2023.1127086)] [Medline: [37223690](https://pubmed.ncbi.nlm.nih.gov/37223690/)]
48. Bessa X, Vidal J, Balboa JC, et al. High accuracy of a blood ctDNA-based multimodal test to detect colorectal cancer. *Ann Oncol*. Dec 2023;34(12):1187-1193. [doi: [10.1016/j.annonc.2023.09.3113](https://doi.org/10.1016/j.annonc.2023.09.3113)] [Medline: [37805131](https://pubmed.ncbi.nlm.nih.gov/37805131/)]
49. Gao Y, Cao D, Li M, et al. Integration of multiomics features for blood-based early detection of colorectal cancer. *Mol Cancer*. Aug 22, 2024;23(1):173. [doi: [10.1186/s12943-024-01959-3](https://doi.org/10.1186/s12943-024-01959-3)] [Medline: [39175001](https://pubmed.ncbi.nlm.nih.gov/39175001/)]
50. Nguyen VTC, Nguyen TH, Doan NNT, et al. Multimodal analysis of methylomics and fragmentomics in plasma cell-free DNA for multi-cancer early detection and localization. *Elife*. Oct 11, 2023;12:RP89083. [doi: [10.7554/eLife.89083](https://doi.org/10.7554/eLife.89083)] [Medline: [37819044](https://pubmed.ncbi.nlm.nih.gov/37819044/)]
51. Liu J, Dai L, Wang Q, et al. Multimodal analysis of cfDNA methylomes for early detecting esophageal squamous cell carcinoma and precancerous lesions. *Nat Commun*. May 2, 2024;15(1):38697989. [doi: [10.1038/s41467-024-47886-1](https://doi.org/10.1038/s41467-024-47886-1)]
52. Liu L, Xiong Y, Zheng Z, et al. AutoCancer as an automated multimodal framework for early cancer detection. *iScience*. Jul 2024;27(7):110183. [doi: [10.1016/j.isci.2024.110183](https://doi.org/10.1016/j.isci.2024.110183)]
53. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. Feb 23, 2009;338(feb23 1):b375. [doi: [10.1136/bmj.b375](https://doi.org/10.1136/bmj.b375)] [Medline: [19237405](https://pubmed.ncbi.nlm.nih.gov/19237405/)]
54. Gui CP, Chen YH, Zhao HW, et al. Multimodal recurrence scoring system for prediction of clear cell renal cell carcinoma outcome: a discovery and validation study. *Lancet Digit Health*. Aug 2023;5(8):e515-e524. [doi: [10.1016/S2589-7500\(23\)00095-X](https://doi.org/10.1016/S2589-7500(23)00095-X)] [Medline: [37393162](https://pubmed.ncbi.nlm.nih.gov/37393162/)]
55. Sujit SJ, Aminu M, Karpinets TV, et al. Enhancing NSCLC recurrence prediction with PET/CT habitat imaging, ctDNA, and integrative radiogenomics-blood insights. *Nat Commun*. Nov 2024;15(1):38605064. [doi: [10.1038/s41467-024-47512-0](https://doi.org/10.1038/s41467-024-47512-0)]

56. Hassett MJ, Uno H, Cronin AM, Carroll NM, Hornbrook MC, Ritzwoller D. Detecting lung and colorectal cancer recurrence using structured clinical/administrative data to enable outcomes research and population health management. *Med Care*. Dec 2017;55(12):e88-e98. [doi: [10.1097/MLR.0000000000000404](https://doi.org/10.1097/MLR.0000000000000404)] [Medline: [29135771](https://pubmed.ncbi.nlm.nih.gov/29135771/)]
57. Steyaert S, Qiu YL, Zheng Y, Mukherjee P, Vogel H, Gevaert O. Multimodal deep learning to predict prognosis in adult and pediatric brain tumors. *Commun Med (Lond)*. Mar 29, 2023;3(1):44. [doi: [10.1038/s43856-023-00276-y](https://doi.org/10.1038/s43856-023-00276-y)] [Medline: [36991216](https://pubmed.ncbi.nlm.nih.gov/36991216/)]
58. Guo W, Liang W, Deng Q, Zou X. A multimodal affinity fusion network for predicting the survival of breast cancer patients. *Front Genet*. 2021;12(709027):709027. [doi: [10.3389/fgene.2021.709027](https://doi.org/10.3389/fgene.2021.709027)] [Medline: [34490038](https://pubmed.ncbi.nlm.nih.gov/34490038/)]
59. Schulz S, Woerl AC, Jungmann F, et al. Multimodal deep learning for prognosis prediction in renal cancer. *Front Oncol*. 2021;11(788740):788740. [doi: [10.3389/fonc.2021.788740](https://doi.org/10.3389/fonc.2021.788740)] [Medline: [34900744](https://pubmed.ncbi.nlm.nih.gov/34900744/)]
60. Cheerla A, Gevaert O. Deep learning with multimodal representation for pancancer prognosis prediction. *Bioinformatics*. Jul 15, 2019;35(14):i446-i454. [doi: [10.1093/bioinformatics/btz342](https://doi.org/10.1093/bioinformatics/btz342)] [Medline: [31510656](https://pubmed.ncbi.nlm.nih.gov/31510656/)]
61. Tan K, Huang W, Liu X, Hu J, Dong S. A multi-modal fusion framework based on multi-task correlation learning for cancer prognosis prediction. *Artif Intell Med*. Apr 2022;126(102260):102260. [doi: [10.1016/j.artmed.2022.102260](https://doi.org/10.1016/j.artmed.2022.102260)] [Medline: [35346442](https://pubmed.ncbi.nlm.nih.gov/35346442/)]
62. Saleh GA, Batouty NM, Haggag S, et al. The role of medical image modalities and AI in the early detection, diagnosis and grading of retinal diseases: a survey. *Bioengineering (Basel)*. Aug 4, 2022;9(8):366. [doi: [10.3390/bioengineering9080366](https://doi.org/10.3390/bioengineering9080366)] [Medline: [36004891](https://pubmed.ncbi.nlm.nih.gov/36004891/)]
63. Wang S, He X, Jian Z, et al. Advances and prospects of multi-modal ophthalmic artificial intelligence based on deep learning: a review. *Eye Vis (Lond)*. Oct 1, 2024;11(1):38. [doi: [10.1186/s40662-024-00405-1](https://doi.org/10.1186/s40662-024-00405-1)] [Medline: [39350240](https://pubmed.ncbi.nlm.nih.gov/39350240/)]
64. Mehta P, Petersen CA, Wen JC, et al. Automated detection of glaucoma with interpretable machine learning using clinical data and multimodal retinal images. *Am J Ophthalmol*. Nov 2021;231:154-169. [doi: [10.1016/j.ajo.2021.04.021](https://doi.org/10.1016/j.ajo.2021.04.021)] [Medline: [33945818](https://pubmed.ncbi.nlm.nih.gov/33945818/)]
65. Xiong J, Li F, Song D, et al. Multimodal machine learning using visual fields and peripapillary circular OCT scans in detection of glaucomatous optic neuropathy. *Ophthalmology*. Feb 2022;129(2):171-180. [doi: [10.1016/j.ophtha.2021.07.032](https://doi.org/10.1016/j.ophtha.2021.07.032)] [Medline: [34339778](https://pubmed.ncbi.nlm.nih.gov/34339778/)]
66. Wu J, Fang H, Li F, et al. GAMMA challenge: Glaucoma grAding from Multi-Modality imAges. *Med Image Anal*. Dec 2023;90(102938):102938. [doi: [10.1016/j.media.2023.102938](https://doi.org/10.1016/j.media.2023.102938)] [Medline: [37806020](https://pubmed.ncbi.nlm.nih.gov/37806020/)]
67. Zhou Y, Yang G, Zhou Y, Ding D. Representation, alignment, fusion: a generic transformer-based framework for multi-modal glaucoma recognition. In: Zhao J, editor. Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention. Oct 1, 2023;Springer. 704-713; Vancouver Convention Centre, Canada. [doi: [10.1007/978-3-031-43990-2_66](https://doi.org/10.1007/978-3-031-43990-2_66)]
68. Wang W, Xu Z, Yu W, Zhao J, Yang J. Two-stream CNN with loose pair training for multi-modal AMD categorization. In: He F, editor. Presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention; Oct 10, 2019; Shenzhen, China. [doi: [10.1007/978-3-030-32239-7_18](https://doi.org/10.1007/978-3-030-32239-7_18)]
69. Vaghefi E, Hill S, Kersten HM, Squirrell D. Multimodal retinal image analysis via deep learning for the diagnosis of intermediate dry age-related macular degeneration: a feasibility study. *J Ophthalmol*. 2020;2020(7493419):7493419. [doi: [10.1155/2020/7493419](https://doi.org/10.1155/2020/7493419)] [Medline: [32411434](https://pubmed.ncbi.nlm.nih.gov/32411434/)]
70. Xu Z, Wang W, Yang J, et al. Automated diagnoses of age-related macular degeneration and polypoidal choroidal vasculopathy using bi-modal deep convolutional neural networks. *Br J Ophthalmol*. Apr 2021;105(4):561-566. [doi: [10.1136/bjophthalmol-2020-315817](https://doi.org/10.1136/bjophthalmol-2020-315817)] [Medline: [32499330](https://pubmed.ncbi.nlm.nih.gov/32499330/)]
71. Wang MH, Xing L, Pan Y, et al. AI-based advanced approaches and dry eye disease detection based on multi-source evidence: cases, applications, issues, and future directions. *Big Data Min Anal*. 2024;7(2):445-484. [doi: [10.26599/BDMA.2023.9020024](https://doi.org/10.26599/BDMA.2023.9020024)]
72. He X, Deng Y, Fang L, Peng Q. Multi-modal retinal image classification with modality-specific attention network. *IEEE Trans Med Imaging*. Jun 2021;40(6):1591-1602. [doi: [10.1109/TMI.2021.3059956](https://doi.org/10.1109/TMI.2021.3059956)] [Medline: [33625978](https://pubmed.ncbi.nlm.nih.gov/33625978/)]
73. Hervella ÁS, Rouco J, Novo J, Ortega M. Multimodal image encoding pre-training for diabetic retinopathy grading. *Comput Biol Med*. Apr 2022;143:105302. [doi: [10.1016/j.combiomed.2022.105302](https://doi.org/10.1016/j.combiomed.2022.105302)]
74. Atse YC, Le Boité H, Bonnin S, Cosette D, Deman P, Borderie L. Improved automatic diabetic retinopathy severity classification using deep multimodal fusion of UWF-CFP and OCTA images. Presented at: Ophthalmic Medical Image Analysis: 10th International Workshop, OMIA 2023, Held in Conjunction with MICCAI 2023; Oct 12, 2023; Vancouver, BC, Canada.
75. Li X, Wen X, Shang X, et al. Identification of diabetic retinopathy classification using machine learning algorithms on clinical data and optical coherence tomography angiography. *Eye (Lond)*. Oct 2024;38(14):2813-2821. [doi: [10.1038/s41433-024-03173-3](https://doi.org/10.1038/s41433-024-03173-3)]

76. Yang J, Yang Z, Mao Z, Li B, Zhang B, et al. Bi-modal deep learning for recognizing multiple retinal diseases based on color fundus photos and OCT images. *Invest Ophthalmol Vis Sci*. 2021;62(8). URL: <https://iovs.arvojournals.org/article.aspx?articleid=2773464> [Accessed 2025-08-14]
77. Peng Z, Ma R, Zhang Y, et al. Development and evaluation of multimodal AI for diagnosis and triage of ophthalmic diseases using ChatGPT and anterior segment images: protocol for a two-stage cross-sectional study. *Front Artif Intell*. 2023;6(1323924):1323924. [doi: [10.3389/frai.2023.1323924](https://doi.org/10.3389/frai.2023.1323924)] [Medline: [38145231](https://pubmed.ncbi.nlm.nih.gov/38145231/)]
78. Flammer J, Konieczka K, Bruno RM, Virdis A, Flammer AJ, Taddei S. The eye and the heart. *Eur Heart J*. May 2013;34(17):1270-1278. [doi: [10.1093/eurheartj/ehf023](https://doi.org/10.1093/eurheartj/ehf023)] [Medline: [23401492](https://pubmed.ncbi.nlm.nih.gov/23401492/)]
79. Allon R, Aronov M, Belkin M, Maor E, Shechter M, Fabian ID. Retinal microvascular signs as screening and prognostic factors for cardiac disease: a systematic review of current evidence. *Am J Med*. Jan 2021;134(1):36-47. [doi: [10.1016/j.amjmed.2020.07.013](https://doi.org/10.1016/j.amjmed.2020.07.013)] [Medline: [32861624](https://pubmed.ncbi.nlm.nih.gov/32861624/)]
80. Chua J, Chin CWL, Hong J, et al. Impact of hypertension on retinal capillary microvasculature using optical coherence tomographic angiography. *J Hypertens*. Mar 2019;37(3):572-580. [doi: [10.1097/HJH.0000000000001916](https://doi.org/10.1097/HJH.0000000000001916)] [Medline: [30113530](https://pubmed.ncbi.nlm.nih.gov/30113530/)]
81. Al-Absi HRH, Islam MT, Refaee MA, Chowdhury MEH, Alam T. Cardiovascular disease diagnosis from DXA scan and retinal images using deep learning. *Sensors (Basel)*. Jun 7, 2022;22(12):4310. [doi: [10.3390/s22124310](https://doi.org/10.3390/s22124310)] [Medline: [35746092](https://pubmed.ncbi.nlm.nih.gov/35746092/)]
82. Lee YC, Cha J, Shim I, et al. Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. *NPJ Digit Med*. Feb 2023;6(1):36732671. [doi: [10.1038/s41746-023-00748-4](https://doi.org/10.1038/s41746-023-00748-4)]
83. Sedlakova J, Daniore P, Horn Wintsch A, et al. Challenges and best practices for digital unstructured data enrichment in health research: a systematic narrative review. *PLOS Digit Health*. Oct 2023;2(10):e0000347. [doi: [10.1371/journal.pdig.0000347](https://doi.org/10.1371/journal.pdig.0000347)] [Medline: [37819910](https://pubmed.ncbi.nlm.nih.gov/37819910/)]
84. Flores JE, Claborne DM, Weller ZD, Webb-Robertson BJM, Waters KM, Bramer LM. Missing data in multi-omics integration: recent advances through artificial intelligence. *Front Artif Intell*. 2023;6(1098308):1098308. [doi: [10.3389/frai.2023.1098308](https://doi.org/10.3389/frai.2023.1098308)] [Medline: [36844425](https://pubmed.ncbi.nlm.nih.gov/36844425/)]
85. Theodos K, Sittig S. Health information privacy laws in the digital age: HIPAA doesn't apply. *Perspect Health Inf Manag*. 2021;18(Winter):11. [Medline: [33633522](https://pubmed.ncbi.nlm.nih.gov/33633522/)]
86. Schwartz PH, Caine K, Alpert SA, Meslin EM, Carroll AE, Tierney WM. Patient preferences in controlling access to their electronic health records: a prospective cohort study in primary care. *J Gen Intern Med*. Jan 2015;30 Suppl 1(Suppl 1):S25-30. [doi: [10.1007/s11606-014-3054-z](https://doi.org/10.1007/s11606-014-3054-z)] [Medline: [25480721](https://pubmed.ncbi.nlm.nih.gov/25480721/)]
87. Amal S, Safarnejad L, Omiye JA, Ghanzouri I, Cabot JH, Ross EG. Use of multi-modal data and machine learning to improve cardiovascular disease care. *Front Cardiovasc Med*. 2022;9(840262):840262. [doi: [10.3389/fcvm.2022.840262](https://doi.org/10.3389/fcvm.2022.840262)] [Medline: [35571171](https://pubmed.ncbi.nlm.nih.gov/35571171/)]
88. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics*. Apr 2016;22(2):303-341. [doi: [10.1007/s11948-015-9652-2](https://doi.org/10.1007/s11948-015-9652-2)] [Medline: [26002496](https://pubmed.ncbi.nlm.nih.gov/26002496/)]
89. Choudhury S, Fishman JR, McGowan ML, Juengst ET. Big data, open science and the brain: lessons learned from genomics. *Front Hum Neurosci*. 2014;8(239):24904347. [doi: [10.3389/fnhum.2014.00239](https://doi.org/10.3389/fnhum.2014.00239)]
90. Shojaei P, Vlahu-Gjorgievska E, Chow YW. Security and privacy of technologies in health information systems: a systematic literature review. *Computers*. 2024;13(2):41. [doi: [10.3390/computers13020041](https://doi.org/10.3390/computers13020041)]
91. Kelly CM, Osorio-Marin J, Kothari N, Hague S, Dever JK. Genetic improvement in cotton fiber elongation can impact yarn quality. *Ind Crops Prod*. Mar 2019;129:1-9. [doi: [10.1016/j.indcrop.2018.11.066](https://doi.org/10.1016/j.indcrop.2018.11.066)]
92. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. *J Med Internet Res*. Nov 1, 2017;19(11):e367. [doi: [10.2196/jmir.8775](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
93. Ahmed SF, Alam MdSB, Hassan M, et al. Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artif Intell Rev*. Nov 2023;56(11):13521-13617. [doi: [10.1007/s10462-023-10466-8](https://doi.org/10.1007/s10462-023-10466-8)] [Medline: [37362885](https://pubmed.ncbi.nlm.nih.gov/37362885/)]
94. Bornstein S. Antidiscriminatory algorithms. *Ala L Rev*. 2018;70(2):519. URL: <https://law.ua.edu/wp-content/uploads/2018/12/4-Bornstein-518-572.pdf> [Accessed 2025-08-14]
95. Miasato A, Reis Silva F. Artificial intelligence as an instrument of discrimination in workforce recruitment. *AUSLEG*. Jan 15, 2020;8(2):191-212. URL: <http://acta.sapientia.ro/acta-legal/legal-main.htm> [Accessed 2025-08-14] [doi: [10.47745/AUSLEG.2019.8.2.04](https://doi.org/10.47745/AUSLEG.2019.8.2.04)]
96. Madan S, Henry T, Dozier J, et al. When and how convolutional neural networks generalize to out-of-distribution category-viewpoint combinations. *Nat Mach Intell*. 2022;4(2):146-153. [doi: [10.1038/s42256-021-00437-5](https://doi.org/10.1038/s42256-021-00437-5)]

97. Sadeghi Z, Alizadehsani R, Cifci MA, et al. A review of explainable artificial intelligence in healthcare. *Computers and Electrical Engineering*. Aug 2024;118:109370. [doi: [10.1016/j.compeleceng.2024.109370](https://doi.org/10.1016/j.compeleceng.2024.109370)]
98. Calaon M, Chen T, Tosello G. Integration of multimodal data and explainable artificial intelligence for root cause analysis in manufacturing processes. *CIRP Annals*. 2024;73(1):365-368. [doi: [10.1016/j.cirp.2024.04.014](https://doi.org/10.1016/j.cirp.2024.04.014)]
99. Rodis N, Sardianos C, Radoglou-Grammatikis P, Sarigiannidis P, Varlamis I, Papadopoulos G. Multimodal explainable artificial intelligence: a comprehensive review of methodological advances and future research directions. *arXiv*. [doi: [10.1109/ACCESS.2024.3467062](https://doi.org/10.1109/ACCESS.2024.3467062)]
100. Zhang X, Shen C, Yuan X, Yan S, Xie L, Wang W, et al. From redundancy to relevance: enhancing explainability in multimodal large language models. *arXiv*. Preprint posted online on 2024.
101. Chen P, Dong W, Wang J, Lu X, Kaymak U, Huang Z. Interpretable clinical prediction via attention-based neural network. *BMC Med Inform Decis Mak*. Jul 9, 2020;20(Suppl 3):131. [doi: [10.1186/s12911-020-1110-7](https://doi.org/10.1186/s12911-020-1110-7)] [Medline: [32646437](https://pubmed.ncbi.nlm.nih.gov/32646437/)]
102. Sharma D, Purushotham S, Reddy CK. MedFuseNet: an attention-based multimodal deep learning model for visual question answering in the medical domain. *Sci Rep*. Oct 6, 2021;11(1):19826. [doi: [10.1038/s41598-021-98390-1](https://doi.org/10.1038/s41598-021-98390-1)] [Medline: [34615894](https://pubmed.ncbi.nlm.nih.gov/34615894/)]
103. Luo B, Teng F, Tang G, et al. StereoMM: a graph fusion model for integrating spatial transcriptomic data and pathological images. *Brief Bioinform*. May 1, 2025;26(3):bbaf210. [doi: [10.1093/bib/bbaf210](https://doi.org/10.1093/bib/bbaf210)] [Medline: [40407386](https://pubmed.ncbi.nlm.nih.gov/40407386/)]
104. Jain S, Wallace BC, editors. *Attention Is Not Explanation*. North American Chapter of the Association for Computational Linguistics; 2019.
105. Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. *Neurocomputing*. Sep 2021;452:48-62. [doi: [10.1016/j.neucom.2021.03.091](https://doi.org/10.1016/j.neucom.2021.03.091)]
106. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: visual explanations from deep networks via gradient-based localization. In: Batra D, editor. Presented at: 2017 IEEE International Conference on Computer Vision (ICCV); Sep 10, 2021; Venice. [doi: [10.1109/ICCV.2017.74](https://doi.org/10.1109/ICCV.2017.74)]
107. Zhang Y, Hong D, McClement D, Oladosu O, Pridham G, Slaney G. Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging. *J Neurosci Methods*. Apr 1, 2021;353(109098):109098. [doi: [10.1016/j.jneumeth.2021.109098](https://doi.org/10.1016/j.jneumeth.2021.109098)] [Medline: [33582174](https://pubmed.ncbi.nlm.nih.gov/33582174/)]
108. Zhang H, Ogasawara K. Grad-CAM-based explainable artificial intelligence related to medical text processing. *Bioengineering (Basel)*. Sep 10, 2023;10(9):1070. [doi: [10.3390/bioengineering10091070](https://doi.org/10.3390/bioengineering10091070)] [Medline: [37760173](https://pubmed.ncbi.nlm.nih.gov/37760173/)]
109. Zambrano Chaves JM, Wentland AL, Desai AD, et al. Opportunistic assessment of ischemic heart disease risk using abdominopelvic computed tomography and medical record data: a multimodal explainable artificial intelligence approach. *Sci Rep*. Nov 29, 2023;13(1):21034. [doi: [10.1038/s41598-023-47895-y](https://doi.org/10.1038/s41598-023-47895-y)] [Medline: [38030716](https://pubmed.ncbi.nlm.nih.gov/38030716/)]
110. Zhao J, Feng Q, Wu P, et al. Learning from longitudinal data in electronic health record and genetic data to improve cardiovascular event prediction. *Sci Rep*. Jan 24, 2019;9(1):30679510. [doi: [10.1038/s41598-018-36745-x](https://doi.org/10.1038/s41598-018-36745-x)]
111. Zhang H, Wang X, Liu C, et al. Detection of coronary artery disease using multi-modal feature fusion and hybrid feature selection. *Physiol Meas*. Nov 1, 2020;41(11):115007. [doi: [10.1088/1361-6579/abc323](https://doi.org/10.1088/1361-6579/abc323)]
112. von Spiczak J, Mannil M, Model H, et al. Multimodal multiparametric three-dimensional image fusion in coronary artery disease: combining the best of two worlds. *Radiol Cardiothorac Imaging*. Apr 2020;2(2):e190116. [doi: [10.1148/ryct.2020190116](https://doi.org/10.1148/ryct.2020190116)] [Medline: [33778554](https://pubmed.ncbi.nlm.nih.gov/33778554/)]
113. Flores AM, Schuler A, Eberhard AV, et al. Unsupervised learning for automated detection of coronary artery disease subgroups. *J Am Heart Assoc*. Dec 7, 2021;10(23):e021976. [doi: [10.1161/JAHA.121.021976](https://doi.org/10.1161/JAHA.121.021976)] [Medline: [34845917](https://pubmed.ncbi.nlm.nih.gov/34845917/)]
114. Ali F, El-Sappagh S, Islam SMR, et al. A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Information Fusion*. Nov 2020;63:208-222. [doi: [10.1016/j.inffus.2020.06.008](https://doi.org/10.1016/j.inffus.2020.06.008)]
115. Qiu S, Miller MI, Joshi PS, et al. Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat Commun*. Jun 20, 2022;13(1):35725739. [doi: [10.1038/s41467-022-31037-5](https://doi.org/10.1038/s41467-022-31037-5)]
116. Gabitto MI, Travaglini KJ, Rachleff VM, et al. Integrated multimodal cell atlas of Alzheimer's disease. *Res Sq*. May 23, 2023;37292694. [doi: [10.21203/rs.3.rs-2921860/v1](https://doi.org/10.21203/rs.3.rs-2921860/v1)] [Medline: [37292694](https://pubmed.ncbi.nlm.nih.gov/37292694/)]
117. Makarious MB, Leonard HL, Vitale D, et al. Multi-modality machine learning predicting Parkinson's disease. *NPJ Parkinsons Dis*. Apr 1, 2022;8(1):35. [doi: [10.1038/s41531-022-00288-w](https://doi.org/10.1038/s41531-022-00288-w)] [Medline: [35365675](https://pubmed.ncbi.nlm.nih.gov/35365675/)]
118. Zhang K, Lincoln JA, Jiang X, Bernstam EV, Shams S. Predicting multiple sclerosis severity with multimodal deep neural networks. *BMC Med Inform Decis Mak*. Nov 9, 2023;23(1):255. [doi: [10.1186/s12911-023-02354-6](https://doi.org/10.1186/s12911-023-02354-6)] [Medline: [37946182](https://pubmed.ncbi.nlm.nih.gov/37946182/)]
119. Ding JE, Thao PNM, Peng WC, et al. Large language multimodal models for new-onset type 2 diabetes prediction using five-year cohort electronic health records. *Sci Rep*. Sep 6, 2024;14(1):20774. [doi: [10.1038/s41598-024-71020-2](https://doi.org/10.1038/s41598-024-71020-2)] [Medline: [39237580](https://pubmed.ncbi.nlm.nih.gov/39237580/)]
120. Bhatt RR, Todorov S, Sood R, et al. Integrated multi-modal brain signatures predict sex-specific obesity status. *Brain Commun*. 2023;5(2):fcad098. [doi: [10.1093/braincomms/fcad098](https://doi.org/10.1093/braincomms/fcad098)] [Medline: [37091587](https://pubmed.ncbi.nlm.nih.gov/37091587/)]

121. Lafci B, Hadjihambi A, Determann M, et al. Multimodal assessment of non-alcoholic fatty liver disease with transmission-reflection optoacoustic ultrasound. *Theranostics*. 2023;13(12):4217-4228. [doi: [10.7150/thno.78548](https://doi.org/10.7150/thno.78548)] [Medline: [37554280](https://pubmed.ncbi.nlm.nih.gov/37554280/)]
122. Liu X, Pan Y, Zhang X, et al. A deep learning model for classification of parotid neoplasms based on multimodal magnetic resonance image sequences. *Laryngoscope*. Feb 2023;133(2):327-335. [doi: [10.1002/lary.30154](https://doi.org/10.1002/lary.30154)] [Medline: [35575610](https://pubmed.ncbi.nlm.nih.gov/35575610/)]
123. Choi Y, Bang J, Kim SY, Seo M, Jang J. Deep learning-based multimodal segmentation of oropharyngeal squamous cell carcinoma on CT and MRI using self-configuring nnU-Net. *Eur Radiol*. Aug 2024;34(8):5389-5400. [doi: [10.1007/s00330-024-10585-y](https://doi.org/10.1007/s00330-024-10585-y)] [Medline: [38243135](https://pubmed.ncbi.nlm.nih.gov/38243135/)]
124. Sundgaard JV, Hannemose MR, Laugesen S, et al. Multi-modal deep learning for joint prediction of otitis media and diagnostic difficulty. *Laryngoscope Investig Otolaryngol*. Feb 2024;9(1):e1199. [doi: [10.1002/lio2.1199](https://doi.org/10.1002/lio2.1199)] [Medline: [38362190](https://pubmed.ncbi.nlm.nih.gov/38362190/)]
125. Callejón-Leblic MA, Blanco-Trejo S, Villarreal-Garza B, et al. A multimodal database for the collection of interdisciplinary audiological research data in Spain. *Auditio*. Sep 2024;8:e109. [doi: [10.51445/sja.auditio.vol8.2024.109](https://doi.org/10.51445/sja.auditio.vol8.2024.109)]
126. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature New Biol*. Aug 3, 2023;620(7972):172-180. [doi: [10.1038/s41586-023-06291-2](https://doi.org/10.1038/s41586-023-06291-2)]
127. Huang D, Yan C, Li Q, Peng X. From large language models to large multimodal models: a literature review. *Appl Sci (Basel)*. 2024;14(12):5068. [doi: [10.3390/app14125068](https://doi.org/10.3390/app14125068)]
128. Qi S, Cao Z, Rao J, Wang L, Xiao J, Wang X. What is the limitation of multimodal LLMs? A deeper look into multimodal LLMs through prompt probing. *Inf Process Manag*. Nov 2023;60(6):103510. [doi: [10.1016/j.ipm.2023.103510](https://doi.org/10.1016/j.ipm.2023.103510)]
129. Liu F, Zhu T, Wu X, et al. A medical multimodal large language model for future pandemics. *NPJ Digit Med*. Dec 2, 2023;6(1):38042919. [doi: [10.1038/s41746-023-00952-2](https://doi.org/10.1038/s41746-023-00952-2)]
130. Xu P, Zhu X, Clifton DA. Multimodal learning with transformers: a survey. *IEEE Trans Pattern Anal Mach Intell*. Oct 2023;45(10):12113-12132. [doi: [10.1109/TPAMI.2023.3275156](https://doi.org/10.1109/TPAMI.2023.3275156)] [Medline: [37167049](https://pubmed.ncbi.nlm.nih.gov/37167049/)]
131. Zhou HY, Yu Y, Wang C, et al. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nat Biomed Eng*. Jun 2023;7(6):743-755. [doi: [10.1038/s41551-023-01045-x](https://doi.org/10.1038/s41551-023-01045-x)]
132. Steurer B, Vanhaelen Q, Zhavoronkov A. Multimodal transformers and their applications in drug target discovery for aging and age-related diseases. *J Gerontol A Biol Sci Med Sci*. Sep 1, 2024;79(9):39126345. [doi: [10.1093/gerona/glae006](https://doi.org/10.1093/gerona/glae006)] [Medline: [39126345](https://pubmed.ncbi.nlm.nih.gov/39126345/)]
133. Takagi Y, Hashimoto N, Masuda H, et al. Transformer-based personalized attention mechanism for medical images with clinical records. *J Pathol Inform*. 2023;14(100185):100185. [doi: [10.1016/j.jpi.2022.100185](https://doi.org/10.1016/j.jpi.2022.100185)]
134. Narhi-Martinez W, Dube B, Golomb JD. Attention as a multi-level system of weights and balances. *Wiley Interdiscip Rev Cogn Sci*. Jan 2023;14(1):e1633. [doi: [10.1002/wcs.1633](https://doi.org/10.1002/wcs.1633)] [Medline: [36317275](https://pubmed.ncbi.nlm.nih.gov/36317275/)]
135. Sha Y, Wang MD. Interpretable predictions of clinical outcomes with an attention-based recurrent neural network. *ACM BCB*. Aug 2017;2017:233-240. [doi: [10.1145/3107411.3107445](https://doi.org/10.1145/3107411.3107445)] [Medline: [32577628](https://pubmed.ncbi.nlm.nih.gov/32577628/)]

Abbreviations

AI: artificial intelligence
AMD: age-related macular degeneration
AUC: area under the curve
CT: computed tomography
ctDNA: circulating tumor DNA
CVD: cardiovascular disease
DL: deep learning
EHR: electronic health record
Grad-CAM: Gradient-weighted Class Activation Mapping
LLM: large language model
MRI: magnetic resonance imaging
NSCLC: non-small cell lung cancer
OCT: optical coherence tomography
TME: tumor microenvironment
XAI: explainable artificial intelligence

Edited by Naomi Cahill; peer-reviewed by Chidinma Madu, Emmanuel Oluwagbade, Victoria Ajibade; submitted 26.04.2025; final revised version received 10.06.2025; accepted 27.06.2025; published 21.08.2025

Please cite as:

Hao Y, Cheng C, Li J, Li H, Di X, Zeng X, Jin S, Han X, Liu C, Wang Q, Luo B, Zeng X, Li K
Multimodal Integration in Health Care: Development With Applications in Disease Management
J Med Internet Res 2025;27:e76557

URL: <https://www.jmir.org/2025/1/e76557>

doi: [10.2196/76557](https://doi.org/10.2196/76557)

© Yan Hao, Chao Cheng, Juanjuan Li, Hongwen Li, Xingsi Di, Xiaoxia Zeng, Shoumei Jin, Xiaodong Han, Chongsong Liu, Qianqian Wang, Bingying Luo, Xianhai Zeng, Ke Li. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 21.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.