REVIEW ARTICLE

Integration of Multimodal Large Language Models in Medical Imaging and Omics Data: A Comprehensive Review

Raja Vavekanand *

Datalink Research and Technology Lab, Islamkot 69240, Sindh, Pakistan

*Corresponding Author: Raja Vavekanand
Email: bharwanivk@outlook.com

Received: 13 November 2024 / Accepted: 17 August 2025

Abstract

Purpose: This review focuses on how Multimodal Large Language Models (MLLMs) and multimodal AI models are advancing healthcare by integrating medical imaging and omics data. By integrating imaging techniques such as MRI, CT, and PET with genomics, transcriptomics, and proteomics, these models offer a comprehensive understanding of diseases, particularly in areas like cancer diagnosis and treatment. The study also highlights the challenges of managing complex datasets and ensuring effective feature selection.

Materials and Methods: Analysed studies leveraging advanced AI models, such as Convolutional Neural Networks (CNNs) and Multimodal Neural Networks (MM-Nets), to integrate diverse data sources. These models enhance medical imaging with omics data to improve disease prediction and management. Applications reviewed include cancer subtype classification, survival outcome prediction, and precision medicine, with a particular focus on non-invasive diagnostic tools.

Results: The findings underscore the transformative potential of multimodal healthcare. They significantly improve the identification of biomarkers and enable personalized treatment approaches. For instance, models like VGG19-CNN and PAGE-Net demonstrated higher accuracy in predicting cancer-specific outcomes and integrating genomic and imaging data. Moreover, the applications to single-cell analysis and radiomics showcased their ability to uncover molecular-level insights, advancing precision medicine.

Conclusion: represents a breakthrough in healthcare, combining diverse data types to deliver actionable insights for disease management. While challenges such as handling complex datasets and ensuring model transparency remain, ongoing advancements in AI technologies are paving the way for their wider adoption. These models hold immense promise for improving diagnostics, guiding treatment strategies, and enhancing patient outcomes, marking a significant step toward the era of personalized medicine.

Keywords: Multimodal Large Language Models; Medical Imaging; Omics Data; Generative Artificial Intelligence.

DOI: https://doi.org/10.18502/fbt.v13i1.20792

1. Introduction

The integration of Multimodal Large Language Models (MLLMs) and multimodal AI models in the medical domain enhances disease diagnosis and treatment, as seen in applications like cancer subtyping and survival prediction [1-3]. One of the transformations that is happening right now is through confluence with imaging and omics data, which will have a major impact on cancer at the molecular granularity. There is a proliferation of omics data that makes it more feasible to build computational models that integrate different types of data [4]. Despite these advances, biological and medical research is shifting towards integrating electronic health records (EHR) with imaging data using advanced AI models [5]. One of the main strategies to connect medical imaging with multi-omics data is phenotypic information extracted from images that can then be used as annotations for omics data. Multimodal approaches have been shown to be advanced in disease diagnosis, where integrating clinical outcome data, bioimaging, and omics improves diagnostic reliability [6]. Cancer research is one key area of healthcare where multimodal approaches may have a significant impact. Studies have found that there are few integrated multi-omics models available to forecast survival outcomes for liver cancer in a variety of patient populations [7]. Medical imaging can be used in conjunction with multi-omics analysis, as evidenced by the successful applications of AI-driven multi-omics integration in ovarian cancer research [8]. The potential of multimodal in cancer research has been highlighted by methods such as Cancer Integration via Multikernel Learning (CIMLR), which has successfully integrated multi-omic data to identify molecular subtypes within cancer [1].

The convergence of imaging and omics data enhances precision medicine and radiation therapy by enabling non-invasive biomarker identification and optimizing treatment strategies, as demonstrated by radiomics-based AI models improving radiation therapy outcomes [9] and CIMLR identifying molecular subtypes in cancer with high accuracy [1]. Radiomics combined using multimodal imaging has been shown to be useful in identifying non-invasive, cheap, and effective biomarkers especially for radiotherapy [9]. Specifically, recent computational frameworks for the analysis of single-cell omics across multiple modalities further underscore a broader relevance of AI in scRNA-seq studies [10]. Adopting multimodal models that couple imaging with omics is an exciting leap in precision diagnosis, subtypes, and personalized treatment. This is facilitated by sophisticated algorithms that converge to redefine the way overhauling molecular complexities of diseases are managed (Figure 1). Here, we present a review of multimodal models in healthcare, outlining recent breakthroughs, challenges, and future directions in this dynamic field.

This review synthesizes the integration of Multimodal Large Language Models (MLLMs) and multimodal AI in healthcare, focusing on their transformative role in medical imaging and omics data, particularly in cancer diagnosis, precision medicine, and single-cell analysis. Unlike prior works such as Antonelli *et al.* [13], which primarily focus on the technical aspects of imaging-omics integration, and Boehm *et al.* [14], which emphasize precision oncology, this review bridges established multimodal AI models like VGG19-CNN [11], PAGE-Net [13], and MM-Net [11] with emerging MLLMs such as Med-ViLLM [14], ChatGPT-4 [11], and LLaMA 3.1 [13]. It evaluates their potential and limitations within clinical contexts. The review introduces a structured glossary to standardize the term "multimodal," addressing the inconsistencies in its usage across the literature. Furthermore, it offers a comparative analysis of model architectures and their performance, while discussing key challenges such as data management complexity and model transparency, and providing actionable recommendations for clinical adoption. This work fills a gap by integrating technical advancements with practical applications, providing a comprehensive resource for researchers and clinicians working towards the adoption of multimodal AI in personalized medicine.

2. Background Overview

The integration of artificial intelligence is revolutionizing healthcare and biomedical research. It enhances medical imaging for accurate diagnoses and streamlines pathology to identify unseen patterns. It unlocks genomics and proteomics secrets, aiding

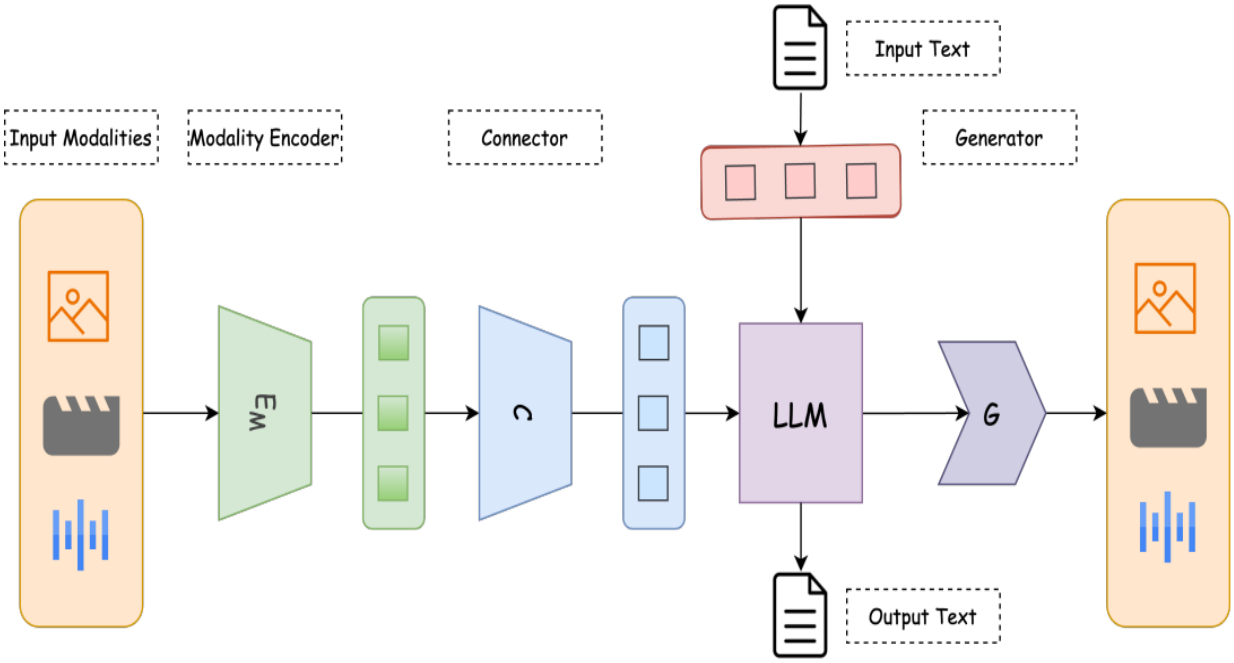

Figure 1. Architecture of a Multimodal Large Language Model, such as Med-ViLLM, illustrating the integration of medical imaging, omics data, and textual data. The model employs modality-specific encoders, Vision Transformer for imaging, feature extractors for omics, and a transformer-based core to align and process multimodal inputs for tasks like cancer diagnosis and precision medicine [11]. Arrows indicate data flow through preprocessing, feature extraction, and fusion layers, enabling cross-modal reasoning and generative outputs [12]

disease understanding. This fusion enables data-driven decisions, driving breakthroughs in diagnosis, treatment, and prevention. Combining human expertise with AI's analytical power. AI's core role in medical imaging, pathology, genomics, and proteomics paves the way for precision medicine, improving patient care and outcomes in unprecedented ways. This glossary addresses the inconsistent use of 'multimodal' in prior reviews [11, 13], providing clarity for researchers and clinicians. The integration of medical imaging and omics data, enabled by MLLMs and multimodal AI, requires a foundational understanding of these data types and their analytical frameworks. This section provides an overview of medical imaging and omics data, highlighting their roles in disease characterization and how advanced AI methods, discussed in later sections, leverage these data for clinical applications.

2.1. Defining Multimodal Large Language Models

Multimodal Large Language Models represent a new generation of AI models that integrate large-scale language processing with the ability to handle diverse data modalities, such as medical imaging and omics data. Unlike standard Large Language Models, which are limited to textual inputs and tasks like medical Q&A or literature summarization, MLLMs employ transformer-based architectures or hybrid frameworks to align and process heterogeneous data, enabling tasks such as cancer subtype classification, survival prediction, and precision medicine. Compared to traditional multimodal AI models, such as Convolutional Neural Networks (CNNs) or Multimodal Neural Networks (MM-Nets), MLLMs benefit from large-scale pretraining on diverse datasets, offering greater generalizability and generative capabilities (producing diagnostic reports). For instance, models like Med-ViLLM and OMNIGPT combine vision and language processing to integrate histopathological images with genomic data, outperforming task-specific models in complex biomedical applications.

2.2. Medical Imaging

Medical imaging techniques are fundamental in delivering detailed anatomical and physiological insights. Unlike recent reviews that focus on domain-specific models [11], this work explores the potential of general-purpose MLLMs like ChatGPT-4 and LLaMA 3.1, evaluating their emerging role in multimodal medical applications. Key modalities in this domain include:

- **Computed Tomography (CT)**: A CT scan is a non-invasive imaging technique using X-rays to produce detailed cross-sectional images of internal structures. Mathematical algorithms reconstruct these images, providing a clear visualization of tissues and organs. It offers a precise, 3D view of the body's interior, enabling diagnosis and treatment of various medical conditions without surgical intervention, and enhancing patient care and outcomes [10, 15].
- **Magnetic Resonance Imaging (MRI)**: MRI operates through the use of strong magnetic fields and radiofrequency pulses, generating high-resolution images of soft tissues. By utilizing differences in water content and molecular properties, MRI is adept at capturing fine anatomical details, especially in soft tissue visualization [15].
- **Positron Emission Tomography (PET)**: PET imaging involves injecting radiolabeled tracers to visualize metabolic activity within tissues. As tracers decay, PET detects positron emissions, providing functional insights into molecular-level physiological processes. This non-invasive technique offers a detailed view of tissue metabolism, enabling diagnosis and monitoring of various diseases, including cancer and neurological disorders.

The fusion of PET with CT in PET/CT imaging combines detailed structural imaging with metabolic activity (Figure 2), offering a comprehensive view for assessing pathology and disease progression [15]. Hematoxylin and Eosin (H&E) staining is a common technique in histology; eosin gives the cytoplasm and extracellular structures a pink hue, while hematoxylin gives the cell nuclei a blue hue. This staining technique improves microscopic analysis and makes characterizing tissue morphology easier [16]. Histology is digitized through Whole Slide Imaging (WSI), which creates high-resolution digital images from entire slides. This invention facilitates remote computing. A foundational understanding of medical imaging, pathology, genomics, proteomics, computer vision, and machine learning/deep learning applications. It also makes it possible to use machine learning for automated pathology analysis [12].

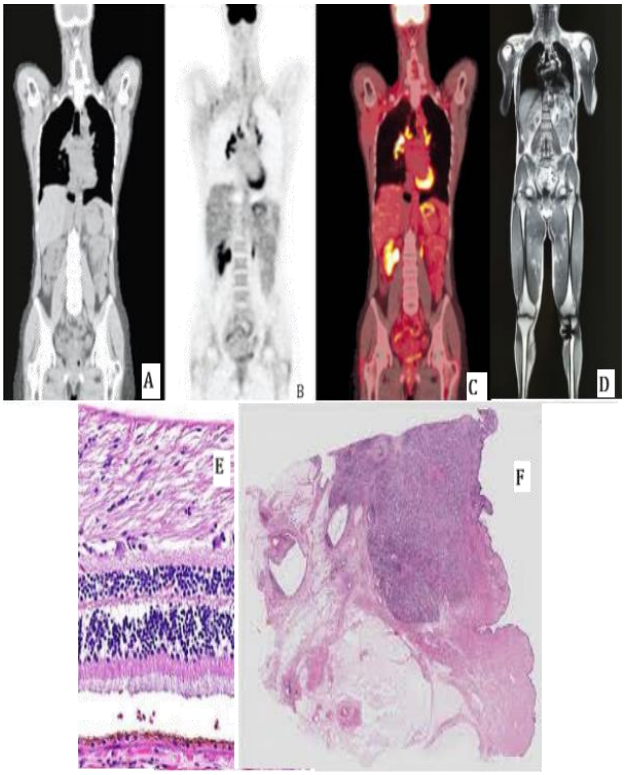

Figure 2. Comparison of different imaging technologies: a: CT [16]. b: PET [16]. d: PET/CT [16]. e: H&E [16]. f: WSI [17]

In medical image analysis, computer vision techniques are game-changers, helping doctors diagnose and interpret images with precision. These innovative tools expertly segment anatomical structures, detect hidden abnormalities, and analyze vital imaging features. Especially in histopathology, machine learning algorithms revolutionize disease diagnosis by identifying subtle patterns in digitized pathology slides. This supports pathologists in making accurate diagnoses, quantifying biomarkers, and predicting patient outcomes. By automating complex tasks, computer vision empowers healthcare professionals to focus on what matters most delivering personalized, life-changing care. Ultimately, this

fusion of technology and medicine saves lives and improves patient outcomes [16].

2.3. Omics Data

The study of molecular biology has advanced through various disciplines that contribute to our understanding of genetic and molecular functions, collectively known as omics:

- **Genomics**: It involves analyzing an organism's entire genetic makeup, including genes, non-coding regions, and structural variations. High-throughput DNA sequencing enables the identification of genetic variations and functional genomic elements, shedding light on the blueprint of life. This knowledge has far-reaching implications for personalized medicine and genetic disease diagnosis [16].
- **Epigenomics**: To explore modifications to DNA and histone proteins that regulate gene expression. Techniques like chromatin immunoprecipitation sequencing (ChIP-seq) reveal DNA-protein interactions, uncovering the intricate epigenetic landscape that influences gene behavior. Understanding epigenomics holds promise for developing novel therapies targeting gene regulation [16].
- **Transcriptomics**: It analyzes RNA transcripts to understand gene expression patterns. RNA sequencing (RNA-Seq) provides a comprehensive view of gene expression, revealing alternative splicing events and outperforming traditional RNA microarrays. This insight enables researchers to understand cellular responses to environmental changes [16].
- **Proteomics**: Systematically examines an organism's complete protein complement. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) enables protein identification and quantification, unraveling cellular processes and shedding light on life's molecular machinery. Proteomics informs our understanding of protein interactions, disease mechanisms, and potential therapeutic targets [16].

These advanced omics techniques, coupled with computer vision, machine learning, and deep learning, are creating a robust framework for exploring molecular and structural aspects of biology. This integration is instrumental in improving diagnostic accuracy, enabling precision medicine, and advancing our understanding of complex diseases (Figure 3). By exploring the intricate relationships between genes, epigenetic modifications, transcripts, and proteins, researchers can develop novel diagnostic tools, therapies, and treatments, ultimately improving human health and well-being.

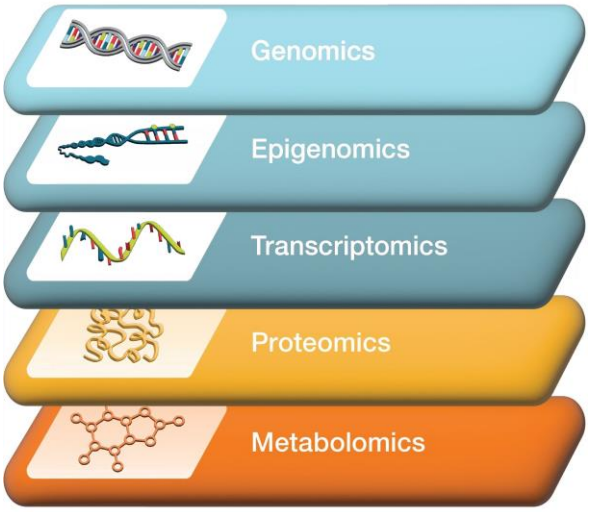

Figure 3. Combining genomics, epigenomics, transcriptomics, and proteomics data enables a comprehensive, systems-level understanding, revealing intricate relationships and insights into biological processes, diseases, and therapeutic targets [18]

3. Data

The data sources and methodologies are central to integrating multimodal data in biomedical research. By analyzing both molecular and imaging datasets, a comprehensive view of how different data types are combined to advance disease characterization and predictive modeling is provided (Table 1).

3.1. Multimodal Data

Modern Biomedical confronts a daunting task integrating large datasets, merging medical imaging and high-throughput omics data, to unlock insights into complex biological systems. The data encompasses information derived from diverse features and sample sets, offering complementary

Table 1. Multimodal Data Sources and Processing Techniques

Data Type	Source	Processing Techniques	Applications
Genomics	National Cancer Institute's Genomic Data Commons (GDC), The Cancer Genome Atlas (TCGA), MOSCATO dataset	High-throughput sequencing, variant analysis	Disease mutation analysis, therapy classification
Epigenomics	Epigenome-Wide Association Studies (EWAS)	DNA methylation profiling, association studies	Epigenetic variation analysis, phenotype association studies
Proteomics	Clinical Proteomic Tumor Analysis Consortium (CPTAC)	Mass spectrometry, protein quantification	Cancer type profiling, protein marker analysis
Transcriptomics	RNA sequencing (RNA-Seq), RNA-microarrays, ENCODE database	RNA-Seq and microarray analysis, cDNA probes for gene expression	Gene expression profiling, transcriptomic characterization
Medical Imaging	TCGA imaging data, Cancer Imaging Archive (TCIA), UK BioBank, Alzheimer's Disease Neuroimaging Initiative (ADNI)	CT, MRI, PET/CT, histopathology staining	Anatomical assessment, disease classification, biomarker tracking
Specialized Methods	NCI Patient-Derived Models Repository (PDMR), Library of Integrated Network-Based Cellular Signatures (LINCS)	Histopathology image tiling, gene expression analysis, digital staining and normalization	Integrative disease modeling, drug response prediction
Histopathology	H&E stained slides, Aperio AT2 scanning, QuPath, Slideflow software	Staining, ROI annotation, tiling and resizing, Reinhard stain normalization	Tumor assessment, image standardization
Transcriptomic Data from PDX	Patient-Derived Xenografts (PDX)	RNA-Seq, log2 transformation, TPM normalization	Transcriptomic changes analysis

insights essential for analyzing biological samples, events, and systems. For the purposes of this review, the integration of at least two data categories, which may include anatomical pathology, genomics, epigenomics, transcriptomics, proteomics, and medical imaging [19]. While these categories serve as a primary focus, multimodal AI applications extend across broader data modalities [20]. The integration of multimodal data is vital for enabling data-driven analyses to address feature selection, classification, regression, unsupervised learning, and association studies, as well as to facilitate predictive model construction for disease detection and classification.

3.2. Molecular Data

This review focuses on omics data generated through high-throughput techniques, specifically genomics, epigenomics, transcriptomics, and proteomics. These omics categories are pivotal in studies aiming to integrate molecular data for disease characterization and predictive modeling.

- **Genomics and Epigenomics**: For genomic data, researchers commonly use datasets from the National Cancer Institute's Genomic Data Commons (GDC) [21]. Cancer Genome Atlas Program (TCGA) [13, 22, 23] MOSCATO dataset [19] employed for training machine learning models, providing extensive data on tumor types, grades, therapy classes, and genomic composition [23]. To incorporate epigenomic data, the studies utilize epigenome-wide association studies (EWAS), which analyze quantifiable epigenetic markers, like DNA methylation, to establish associations between epigenetic variation and phenotypes [24, 25].

- **Proteomics**: The Clinical Proteomic Tumor Analysis Consortium (CPTAC) [26], offers a public catalog of proteomic and genomic data across various cancer types, with significant overlap with the TCGA. The combination of CPTAC and TCGA data provides a detailed molecular profile of patient disease states [14].
- **Transcriptomics**: The data used in these studies is often derived through RNA sequencing (RNA-Seq) [27] or RNA-microarrays [28], focusing on mRNA levels for predictive modeling. RNA-Seq provides an unbiased, high-throughput view of the transcriptome, while RNA-microarrays utilize complementary DNA (cDNA) probes to measure gene expression. Transcriptomics data is frequently sourced from the ENCODE database for studies in both epigenomics and transcriptomics [29].

3.3. Medical Imaging Data

The integration of imaging and molecular data is crucial in biomedical research for a comprehensive understanding of health and disease. Approaches vary, with some studies exclusively leveraging the TCGA database for both molecular and imaging data, while others utilize additional imaging databases.

- **TCGA-Based Imaging Data**: The TCGA database houses diverse imaging data, including CT, MRI, PET, and PET/CT scans, covering healthy controls and various cancer types [13]. This dual repository allows researchers to study imaging and molecular data within a single, integrated framework.
- **Extended Imaging Databases**: Some studies expand beyond TCGA, integrating data from sources such as private clinical cancer institutions and the Cancer Imaging Archive (TCIA) [13, 22, 30]. TCIA provides images from various techniques, capturing a range of scenarios from healthy controls to different cancer types, UK BioBank [31, 32]. Offers a wealth of body and cardiac imaging, genetic data, lifestyle measures, biological phenotyping, and health records. Neuroimaging studies often utilize datasets such as Alzheimer's Disease Neuroimaging Initiative (ADNI) [31], and the Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) [32]. While ADNI limits open access to its data, ENIGMA facilitates free, standardized data analysis, enabling collaborators to independently process their data [16].
- **Specialized Methodologies**: In a unique approach, Partin *et al.* [33] trained models on data type. Data collection comprised multiple stages, integrating drug descriptors, gene expression profiles, and histology image tiles into a unified dataset.
- **Histopathology and Omics Characterization**: Initial data were sourced from the NCI Patient-Derived Models Repository (PDMR) [34], which provides histopathology assessments, whole-exome sequencing, and RNA-Seq for tumor subsets. This preliminary step established a baseline for histological and omic data characterization.
- **Transcriptomic Data from Patient-Derived Xenografts (PDX)**: Transcriptomic data were gathered from PDX models through RNA-Seq, then transformed into Transcripts Per Kilobase million (TPM). Further log2 transformation and standardization were performed, zero mean and unit standard deviation. The Library of Integrated Network-Based Cellular Signatures (LINCS) project provided landmark genes to capture significant transcriptomic changes [35].
- **Histopathology and Imaging Processing**: Histopathology slides were digitized at 20x magnification using an Aperio AT2 scanner [36]. A board-certified pathologist verified consistency with original diagnoses and annotated regions of interest (ROIs) using QuPath [36]. Whole slide images were processed into 299x299 pixel tiles (302 µm x 302 µm) via Slide flow software [37]. Background filtering (60% threshold) and digital stain normalization using the Reinhard method ensured uniformity, standardizing tiles to mean 0 and unit variance for analysis [38].

This extensive collection and processing of multimodal data underscore the transformative potential of combining molecular and imaging data for disease prediction and characterization.

4. Methods

Specifically addressing statistical and machine learning techniques, these methods are categorized based on their application to supervised classification,

regression, clustering, and network analysis (Figure 4), highlighting significant contributions within the field, such as Antonelli *et al.*'s foundational survey on multimodal data integration [13].

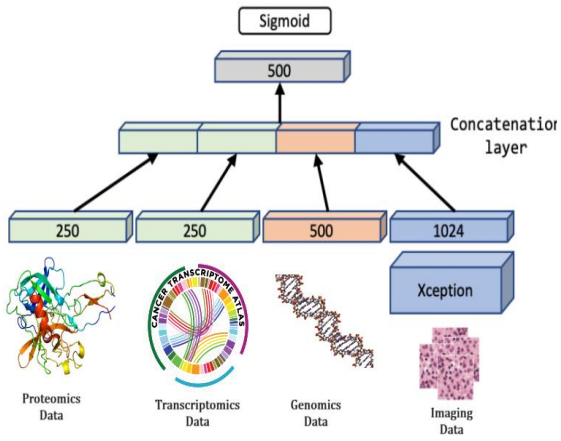

Figure 4. Multimodal Network (MM-Net) illustration, adapted from Partin *et al.*, showcasing integrated analysis of diverse data types [33]

4.1. Classification and Regression Techniques

In supervised classification, numerous algorithms have been developed to handle the inherent complexity of biomedical data, which is often characterized by variability and noise. Supervised classification tasks in biomedical contexts involve functions that assign real-valued labels to data. Techniques like regression analysis are especially useful in these scenarios, where a regularization term controls the number of independent variables involved. For instance, the Lasso method introduces an L1-penalty term to the objective function, effectively handling cases with highly correlated features [39]. These classification and regression techniques, such as Canonical Correlation Analysis, enable precise disease prediction, which is critical for applications like cancer subtype classification discussed in Section 5.

- **Canonical Correlation Analysis (CCA)**: Identifies co-expressed features across modalities, and has also been adapted for multimodal data integration. Yan *et al.* extended traditional incorporating a penalty term that considers disease status, utilizing Laplacian matrices to account for patient diagnoses, thereby enhancing predictive performance in disease-specific contexts.

4.2. Clustering and Unsupervised Methods

In the analysis of unlabeled data, clustering plays an essential role. Unlike classification or regression models, clustering groups similar samples and segregates dissimilar ones without prior labels, offering critical insights into sample relationships and biological themes.

- **lmQCM Clustering**: The lmQCM algorithm allows for overlapping groups, enabling genes to belong to multiple clusters, providing flexibility in representing complex biological relationships.

  Hierarchical Clustering: Employed by Diehn *et al.* [41]. hierarchical clustering identifies functional themes among genes and uncovers modular structures with topological overlaps in networks, which is particularly useful for understanding gene interaction networks in biomedical data.

4.3. Network Analysis and Deep Learning in Imaging and Omics

Network analysis leverages gene expression data to construct networks representing interactions among cell components, facilitating the interpretation of complex biological processes. Recent applications of deep learning include efforts to map tumor gene expression profiles to MRI-based tumor morphology, illustrating how imaging data can enhance omics-based insights. Although deep learning for integrated omics and imaging analyses is still emerging, these approaches hold promise for advancing multimodal data integration [16].

4.4. Model Description

Bioinformatics studies are increasingly focusing on synergistic models that integrate imaging and omics data, moving beyond traditional single-modality approaches. Review several pivotal studies (Figure 5), emphasizing innovations in multimodal AI applications [42].

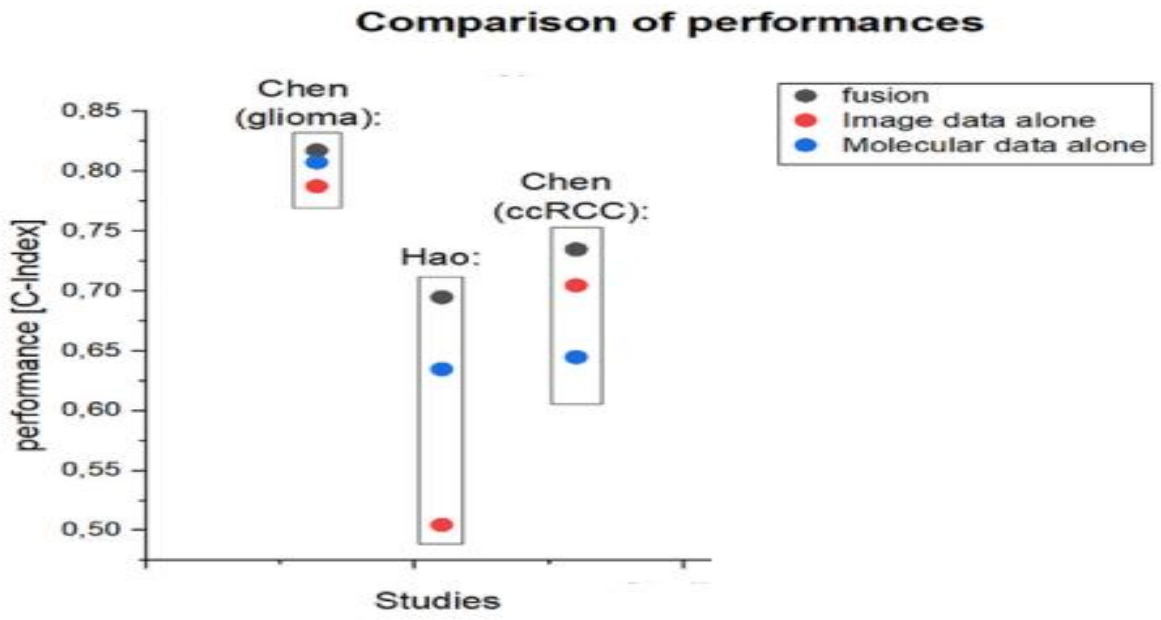

Figure 5. Comparison of different performances in studies, adapted from Shneider *et al.* [42]

- **Chen *et al.*'s VGG19-CNN Model**: In their study, they utilized a VGG19-based Convolutional Neural Network (CNN) to extract features from regions of interest in Whole Slide Imaging (WSI) data [2]. Simultaneously, self-normalizing networks processed omics data, enhancing feature extraction and minimizing overfitting risks. This strategic integration of image and omics data produced a multidimensional array that enriched the model's predictive capacity, marking a significant advancement over traditional single-modality approaches.

  Hao *et al.*'s PAGE-NET Model: Developed the PAGE-NET model, an innovative approach that combines genomic data, patient age, and H&E-stained images. The model uses a texture-based CNN that incorporates a patch aggregation strategy to capture the nuances across data modalities. Notably, PAGE-NET has been effective in predicting Cancer-Specific Survival (CSS) within a Cox proportional hazards model framework, showcasing its versatility and efficacy in clinical outcome prediction [3].

4.5. Multimodal Neural Networks (MM-Net)

The multimodal neural network (MM-Net), designed to predict drug response in patient-derived xenografts (PDXs) [33]. MM-Net was evaluated across six distinct models that varied in feature sets and sample types, including:

- **Unimodal Models**: The UME-Net and UMH-Net models focused on specific data types, with UME-Net for gene expression and UMH-Net employing a CNN to analyze histological features. The LGBM model, optimized for gene expression data, used gradient boosting techniques.
- **Integrated Multimodal Model (MM-Net)**: MM-Net combined data from drug descriptors, gene expression, and histology, demonstrating the enhanced predictive power of multimodal data integration. The integration of diverse modalities improved performance metrics, particularly the Matthews Correlation Coefficient (MCC), underscoring the potential of multimodal approaches in achieving robust predictive outcomes.

These methodologies underscore the progress in multimodal data integration, enabling sophisticated, data-driven predictions in biomedical contexts.

Multimodal neural networks, including MLLMs and multimodal AI models, are essential for integrating medical imaging and omics data. Key models include VGG19-CNN, PAGE-Net, and MM-

Net, each with unique architectures, training datasets, and evaluation metrics.

VGG19-CNN: This model uses a 19-layer CNN to extract features from Whole Slide Imaging (WSI) data and integrates omics data via late fusion. Trained on The Cancer Genome Atlas (TCGA), which includes ~10,000 WSI samples and genomic profiles, it achieved an AUC of 0.85 for cancer survival prediction, though its high computational cost limits scalability [2].

PAGE-Net: Employing a patch-based CNN architecture, PAGE-Net integrates H&E-stained images, genomic data, and patient age for cancer-specific survival (CSS) prediction. Trained on ~2,000 breast cancer samples from TCGA, it achieved an MCC of 0.78 for CSS prediction, but its generalizability is limited to breast cancer [3].

MM-Net: Designed for drug response prediction, MM-Net combines drug descriptors, gene expression, and histological images using a hybrid fusion approach. Trained on ~1,500 PDX samples, it achieved an MCC of 0.82 for drug response prediction but requires complex preprocessing and significant computational resources [33].

4.6. Comparative Analysis of Multimodal Models

To elucidate the strengths and limitations of multimodal models, we compare key models such as VGG19-CNN, PAGE-Net, and MM-Net across several dimensions: architecture, data fusion strategy, performance metrics, and clinical applicability. VGG19-CNN, a deep convolutional neural network, excels in extracting features from Whole Slide Imaging (WSI) data but requires significant computational resources and is primarily suited for imaging-heavy tasks [2]. PAGE-Net integrates genomic data, patient age, and H&E-stained images using a patch-based convolutional approach, achieving high interpretability for cancer-specific survival (CSS) prediction (MCC: 0.78), but is limited to specific cancer types [3]. MM-Net, designed for prediction of drug response in patient-derived xenografts, combines drug descriptors, gene expression, and histology through hybrid fusion, offering robust performance (MCC: 0.82) but requiring complex training pipelines [33]. Table 2 summarizes these comparisons, highlighting trade-offs in scalability, generalizability, and clinical utility.

The methodologies discussed in this section, including supervised classification, clustering, and multimodal neural networks like VGG19-CNN and MM-Net, provide the technical foundation for integrating medical imaging and omics data. These approaches enable the extraction, alignment, and analysis of heterogeneous data types, facilitating applications in disease diagnosis, prognosis, and treatment personalisation. The following section examines how these methods are applied in practical healthcare settings, including the classification of

Table 2. Comparison of Multimodal Models

Model	Architecture	Data Modalities	Fusion Strategy	Performance Metrics	Clinical Applicability	Strengths	Limitations
VGG19-CNN	19-layer CNN	WSI, omics	Late fusion	AUC: 0.85 (cancer survival)	Cancer diagnosis, prognosis	Robust image feature extraction	High computational cost, imaging-focused
PAGE-Net	Patch-based CNN	H&E images, genomics, age	Patch-based fusion	MCC: 0.78 (CSS prediction)	Cancer-specific survival prediction	Interpretable texture analysis	Limited to specific cancer types
MM-Net	Multimodal neural network	Drug descriptors, gene expression, histology	Hybrid fusion	MCC: 0.82 (drug response)	Drug response prediction in PDX models	Comprehensive multimodal integration	Complex training, data preprocessing

cancer subtypes, precision medicine, and single-cell analysis. It discusses their impact on clinical outcomes and the challenges associated with their implementation.

Table 3 compares key multimodal models, highlighting their data types, outcomes, and clinical relevance. Multimodal AI models like VGG19-CNN and PAGE-Net, which process imaging and omics data, excel in specific tasks like cancer survival prediction (AUC: 0.85 [2]) and cancer-specific survival prediction (MCC: 0.78 [3]), but their designs limit generalizability. In contrast, MLLMs like Med-ViLLM and BioMedLM, which integrate text, imaging, and omics, offer more flexibility for tasks like multi-cancer diagnostics (AUC: 0.87 [11]) and disease pattern analysis (AUC: 0.84 [14]). Emerging models like ChatGPT-4 show potential for diagnostic support but lack clinical validation [11]. This table highlights the balance between task-specific accuracy and broader applicability for clinical use.

5. Overview of Multimodal in Medical Imaging

Recent advancements in multimodal large language models have unlocked new potential in healthcare (Table 4). Integration has proven especially valuable in biological and medical research, allowing for a more comprehensive understanding of the human body and disease mechanisms. By merging data from medical images with genetic and molecular insights (Figure 6), multimodal can provide a holistic view of health and disease, with applications ranging from cancer prognosis to precision medicine [4, 5].

The evaluation of Multimodal Large Language Models (MLLMs) in medical imaging and omics data integration is based on several key criteria: (1) Performance Metrics, including accuracy, area under the curve (AUC), Matthews correlation coefficient (MCC), sensitivity, and specificity, as reported in empirical studies for tasks like cancer diagnosis and survival prediction; (2) Generalizability, assessing the model's ability to perform across diverse datasets,

Table 3. Comparative Summary of Key Multimodal Models Comparative summary of key Multimodal Large Language Models (MLLMs) and multimodal AI models, detailing data types, outcomes, accuracy metrics, and clinical relevance. This table highlights the strengths and limitations of each model, guiding their application in healthcare

Model	Type	Data Type	Outcome	Accuracy Metrics	Clinical Relevance	Source
VGG19-CNN	Multimodal AI	WSI, gene expression	Cancer survival prediction	AUC: 0.85	Prognostic assessment in cancer (e.g., breast, lung)	[2]
PAGE-Net	Multimodal AI	H&E images, genomics, patient age	Cancer-specific survival (CSS) prediction	MCC: 0.78	Breast cancer prognosis, risk stratification	[3]
MM-Net	Multimodal AI	Drug descriptors, gene expression, histology	Drug response prediction in PDXs	MCC: 0.82	Personalized treatment in cancer therapy	[33]
Med-ViLLM	MLLM	Text, imaging (MRI, CT, WSI), omics	Disease diagnosis, prognostic prediction	AUC: 0.87	Precision medicine, multi-cancer diagnostics	[11]
BioMedLM	MLLM	Text, genomics, imaging (CT, PET)	Disease pattern analysis	AUC: 0.84	Early detection of cancer and rare diseases	[11, 14]
ChatGPT-4	MLLM (emerging)	Text, imaging (emerging), clinical data	Medical Q&A, diagnostic support	Not fully validated	Potential for patient triage, diagnostics	[11]

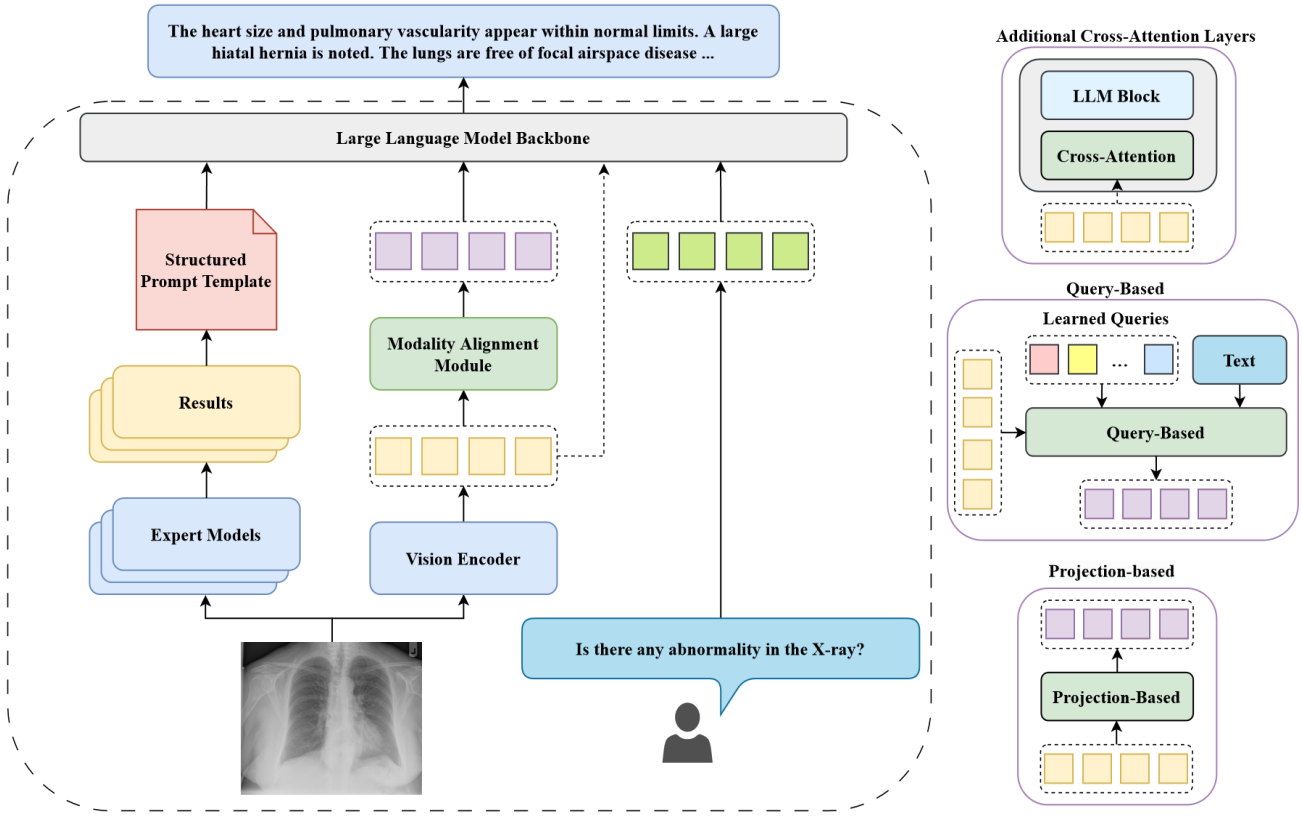

Figure 6. MLLMs consist of core modules and pipelines with modality alignment modules, including expert-model-based prompt augmentation methods [11]

including rare or complex cases; (3) Computational Efficiency, evaluating the resources required for training and inference, critical for clinical scalability; (4) Interpretability, measuring the transparency of model predictions, particularly for genomic and imaging-based outputs; and (5) Clinical Applicability, determining the model's readiness for real-world medical settings based on validation studies and regulatory considerations. Limitations presented in Tables 3 and 4 are derived from empirical studies [2, 3, 13, 14, 33, 43], expert reviews of multimodal AI challenges [11], and the authors' synthesis of recurring issues in the literature, such as data diversity and explainability constraints.

5.1. Applications in Cancer Research

Cancer research is a major focus as scientists work to enhance survival predictions and improve treatment precision. Models like Cancer Integration via Multikernel Learning (CIMLR) enable the integration of diverse data types to identify molecular subtypes of cancer, enhancing our understanding and approach to cancer treatment [44]. Multimodal has shown notable success in predicting survival outcomes for liver cancer patients and classifying ovarian cancer subtypes, underscoring its transformative potential in oncology [6, 7]. Figure 5 illustrates the workflow for integrating medical imaging and omics data in cancer diagnosis using MLLMs and multimodal AI models. It shows the pipeline from data preprocessing (e.g., genomic data normalization, histopathological image segmentation) to feature extraction (e.g., Vision Transformers for Med-ViLLM [11] or CNNs for PAGE-Net [3]) and multimodal fusion for predicting cancer subtypes or survival. PAGE-Net combines H&E-stained images, genomic data, and patient age, achieving an MCC of 0.78 for cancer survival prediction [3], while Med-ViLLM integrates imaging, omics, and clinical reports to enhance diagnostic accuracy [11]. This workflow highlights the benefits of combining imaging and omics data for precision oncology but also addresses challenges like data heterogeneity and computational complexity.

Table 4. Summarizing the Multimodal Large Language Models in medical imaging and omics data highlighting their names, tasks and limitations

Model	Task	Description
BioMedLM	Integrates genomics, clinical, and imaging data	BioMedLM focuses on synthesizing text with genomics and imaging, analyze multi-source biomedical data for insights into disease patterns and treatment outcomes.
MedPaLM	Medical Q&A and diagnostics from imaging & textual data	Designed for medical Q&A and diagnosis support by interpreting text and imaging, particularly useful for identifying symptoms and aiding in diagnostic workflows.
BioGPT	Summarizes biomedical literature, connects text & gene data	Its generates summaries of biomedical literature and links it with genetic data, offering insights into gene-disease relationships and relevant research.
GatorTron	Summarizes radiology and clinical reports	Provides detail clinical report summaries, emphasizing radiology, to support quick comprehension of patient imaging records.
OMNIGPT	Diagnostics with genomics and imaging integration	Enables diagnostics by combining genomics and imaging, supporting precision medicine through its ability to connect complex multimodal data for deeper insights.
UNITER-med	Aligns histopathology images with textual data	This model aligns image and text data for histopathology analysis, helping pathologists link tissue-level details with clinical findings for better disease characterization.
MedCLIP	Contrastive learning for X-ray and image analysis	Based on CLIP technology, it focuses on enhancing X-ray analysis using contrastive learning to find clinically relevant features across various imaging types.
ImageBERT-med	Disease detection across diverse medical images	This model extends the capabilities to medical images, identifying diseases across imaging modalities, including X-rays, MRIs, and CT scans, improving diagnostic accuracy.
OmicsGPT	Integrates text with multi-omics data (genomics, proteomics)	Combines text with multi-omics data, helping researchers understand complex interactions among genomics, proteomics, and clinical data for disease research.
Med-ViLLM	Disease prediction with text, imaging, and omics data	Vision-language model, enables disease prediction by synthesizing information from multiple modalities, supporting more holistic patient assessments.
ChatGPT-4	Medical analysis, patient Q&A, general diagnostic support	Versatile model that offers patient-centered Q&A, diagnostic support, and medical data interpretation, with expanding potential for multimodal inputs.
LLaMA 3.1	Multimodal integration, large-scale medical knowledge tasks	Its provides scalable support for multimodal data integration, interpreting extensive imaging and genomic datasets, with particular applications in predictive analytics.

5.2. Precision Medicine and Radiation Therapy

Advancing medicine and radiation therapy AI models analyze medical images to identify biomarkers biological indicators that guide personalized treatment plans. This process is not only fast and cost-effective but also non-invasive, making it a valuable tool in tailored patient care [8]. Radiomics, a technique that extracts features from medical images, combined with AI, is being used to optimize radiation therapy, allowing for accurate biomarker identification and improving treatment efficacy [1, 45].

5.3. Innovations in Single-Cell Analysis

Single-cell analysis is another promising area for multimodal, where the combination of imaging and omics data reveals insights into cellular behavior and disease mechanisms. AI-driven models are adept at integrating diverse datasets from single cells, including genetic, protein, and molecular data, providing a clearer view of cell function and disease progression [9].

5.4. Technological Foundations of Multimodal AI

The success of multimodal in medicine is rooted in recent advancements in both imaging and omics technologies. It allows to visualize anatomical

structures and processes with high precision. On the other hand, omics technologies provide comprehensive insights into an organism's genetic and molecular profile. Combining these datasets requires sophisticated methods, such as supervised classification algorithms and unsupervised clustering, which can handle the complexity of multi-source data [10, 36]. The power of MLLMs lies in their ability to leverage large-scale pretraining and transformer architectures, distinguishing them from traditional multimodal AI models like CNNs, which are often task-specific and lack generative capabilities. For instance, while VGG19-CNN excels in feature extraction from imaging data, MLLMs like Med-ViLLM integrate imaging, omics, and text to generate comprehensive diagnostic insights, offering a more holistic approach to precision medicine.

5.5. Key Models and Computational Approaches

Advanced AI models are instrumental in integrating and analyzing multimodal data. Particularly, Convolutional Neural Networks (CNNs) play a pivotal role in interpreting complex biomedical data [46]. The VGG19-CNN and self-normalization have been applied to predict cancer survival rates and classify cancer subtypes [15]. Other innovative models, such as PAGE-NET, leverage patch-wise CNN techniques to integrate genomic data, patient age, and histological images, while MM-Net combines drug descriptors, gene expression, and histological features to predict patient responses to treatments [16].

General-purpose Multimodal Large Language Models (MLLMs) like ChatGPT-4 and LLaMA 3.1 are included in this review due to their emerging multimodal capabilities, which hold promise for biomedical applications despite limited validated clinical use to date. ChatGPT-4, developed by OpenAI, has demonstrated potential in medical Q&A and diagnostic support by processing textual and, increasingly, imaging inputs, with preliminary studies exploring its ability to interpret radiology reports [12]. LLaMA 3.1, developed by Meta AI, offers scalable multimodal integration, with research indicating its adaptability for processing large-scale imaging and genomic datasets [43, 46]. While these models lack extensive validation in clinical settings, their transformer-based architectures and pretraining on diverse datasets suggest potential for tasks like disease prediction and personalized medicine. Ongoing research is exploring fine-tuning these models for specific medical tasks, such as integrating histopathology images with omics data, to enhance their clinical relevance [11]. However, challenges such as limited explainability and the need for domain-specific fine-tuning must be addressed to ensure their safe and effective use in healthcare. Table 5 compares key models used in multimodal integration of medical imaging and omics data. Models like BioMedLM and Med-ViLLM utilize transformer-based architectures to analyze disease patterns and precision medicine, though they struggle with diverse imaging datasets. Convolutional models such as VGG19-CNN and PAGE-Net excel in tasks like cancer survival prediction but face issues with computational efficiency and generalizability. Emerging models like ChatGPT-4 and LLaMA 3.1 show potential for medical Q&A and analytics, but need more validation for clinical use. This overview highlights the strengths and limitations of these models, emphasizing the need for improvements in interpretability and scalability.

Table 5 highlights various scenarios where MLLMs are enhancing diagnostic accuracy, supporting clinical decision-making, and personalizing patient care through the integration of imaging, genomics, and clinical data. The applications range from radiology report generation to omics-based cancer diagnosis, demonstrating the transformative potential of MLLMs in diverse clinical settings.

6. Ethical, Interpretability, and Regulatory Considerations

The integration of Multimodal Large Language Models (MLLMs) and multimodal AI models into healthcare presents significant opportunities but also several ethical, interpretability, and regulatory challenges. One of the primary concerns is bias in training data, as datasets such as The Cancer Genome Atlas (TCGA) and PDXNet often lack sufficient diversity in terms of race, ethnicity, and socioeconomic status [22, 33]. This limited representation can lead to biased predictions, particularly in cancer subtype classification, where models may perform poorly for underrepresented

Table 5. Clinical examples of multimodal large language model applications in healthcare

Clinical Scenario	Application	Impact	Source
Radiology Report Generation and Error Detection	**ChatRadio-Valuer** analyzes chest X-rays to suggest potential errors (e.g., pleural effusion).	Assists in error detection, improving radiology report accuracy, and leading to better patient outcomes.	[47]
Dermatological Diagnosis via Skin Image Analysis	**SkinGPT-4** analyzes skin lesions for actinic keratosis diagnosis and treatment recommendations.	Provides rapid and accurate dermatological diagnoses, assisting clinicians in decision-making.	[48]
Multimodal Analysis for Disease Diagnosis	MLLM integrates imaging, genomics, and clinical assessments to diagnose rare cancers.	Facilitates early detection of complex diseases that may be overlooked using traditional methods.	[49]
Clinical Decision Support in Emergency Settings	MLLM analyzes symptoms, medical history, and vital signs for triaging patients in emergencies.	Supports rapid clinical decision-making in high-pressure environments, improving patient outcomes.	[50]
Omics-Based Diagnosis in Oncology	MLLM analyzes genomic and clinical data to recommend targeted therapies for cancer patients.	Personalizes treatment based on genomic information, advancing precision medicine.	[48, 49]

patient groups [51]. Addressing this bias involves diversifying training datasets, applying fairness-aware algorithms like adversarial debiasing, and validating models across a broader range of populations to ensure equitable outcomes [52].

Another critical issue is patient privacy, especially when handling sensitive data from medical imaging, omics, and clinical reports. Regulations like HIPAA in the U.S. and GDPR in the EU must be adhered to in order to safeguard patient confidentiality. Techniques such as differential privacy and federated learning can help protect patient data while enabling model training, though they introduce additional computational complexities [51-53]. Ensuring that privacy is maintained while still allowing models to learn from large datasets is a delicate balance. Model interpretability remains a major hurdle for clinical adoption. Many MLLMs, like Med-ViLLM and BioMedLM, operate as black-box models, making it difficult for healthcare professionals to understand how predictions are made. This lack of transparency undermines trust and hampers clinical integration. Solutions like SHAP and attention visualization are promising, but these techniques require further development to make multimodal models more interpretable [52]. Regulatory hurdles present challenges in deploying AI models in healthcare. Regulatory bodies such as the FDA and EMA require models to undergo rigorous validation to ensure they are safe, effective, and fair. Models like VGG19-CNN, which have high computational costs, also face challenges in real-time deployment [51]. Moving forward, collaboration with regulatory agencies and developing standardized validation protocols will be essential to streamline AI adoption in healthcare [54].

7. Discussion

Combining imaging and omics data enhances disease classification and treatment prediction by leveraging multimodal data, with studies showing improved prognostic accuracy for cancers like liver and lung using integrated omics and imaging data [55, 56]. By leveraging machine learning and advanced analytics, it addresses complex relationships among biological processes, health indicators, and risk factors, proving versatile in modern healthcare [57]. In oncology, multimodal approaches enhance prognostic accuracy and personalize treatment. AI models, combining omics and histopathological imaging data, outperform unimodal models in predicting outcomes for liver, lung, renal, and breast cancers [55, 56]. For complex diseases like head and neck squamous cell carcinoma and neuroendocrine tumors, it captures intricacies and improves predictive precision [56, 58]. It also predicts therapeutic responses, as seen in rivastigmine treatment for Alzheimer's disease, showcasing potential in personalized treatment

planning [59]. Rapid diagnostic capabilities, observed in necrotizing enterocolitis, demonstrate multimodal strength in handling multifaceted datasets [60].

Advancements in models like ELMO and MORONET integrate large-scale models into biomedical research, enhancing multimodal applications [62, 63]. Attribute reduction algorithms for Alzheimer's disease demonstrate new clinical pathways [63]. It improves prognostic accuracy, with UMAP embedding and CNNs showing promise in multi-omics data integration for cancer prediction and survival analysis [64]. This advancement enables personalized treatment strategies and improved outcomes. Extends to neurological diseases, predicting disease progression and treatment recommendations for conditions like Alzheimer's [65]. In inflammatory bowel disease management, models predicting treatment responses and prognosis, and personalizing treatment strategies [12, 43]. Transformative potential in disease management and patient care is evident, offering substantial promise for advancing disease classification and treatment prediction through comprehensive data analysis and predictive modeling [20].

8. Conclusion

The integration of multimodal LLM in medicine holds remarkable promise for transforming healthcare by advancing disease diagnosis. Combining imaging and omics data allows for a more holistic view of diseases, where it enhances prognostic capabilities and facilitates the identification of molecular subtypes. This integrative approach supports precision medicine and radiation therapy by enabling non-invasive identification and optimizing treatment strategies tailored to individual patient profiles. Despite the progress made, significant challenges persist, including data management complexities and the need for robust feature selection. As multimodal continues to evolve, addressing these challenges will be critical in ensuring its successful implementation in clinical settings (Table 6). The development of ethical guidelines and standards for AI transparency will be essential in gaining clinician and patient trust. Overall, the ongoing advancements in a new era in healthcare with the potential to fundamentally reshape disease management and improve clinical decision-making through personalized medicine.

9. Future Perspectives

The integration of Multimodal Large Language Models (MLLMs) in healthcare holds significant promise, but several challenges must be addressed for their successful deployment and scalability. Scalability remains a primary concern as these models must efficiently handle vast datasets from diverse sources, including medical imaging, genomics, and patient records. As healthcare data grows, leveraging cloud computing and distributed systems will be essential to support these AI models, ensuring they are computationally efficient and scalable across various healthcare settings. Fine-tuning models for clinical settings will also be crucial. While MLLMs offer impressive generalization, there is a need for continuous model adaptation to reflect real-time changes in disease patterns and demographics. Techniques like incremental learning and transfer learning could help these models stay updated with evolving patient data, particularly in underrepresented groups or rare diseases.

As the use of AI in healthcare expands, regulatory frameworks will play a pivotal role. Compliance with data privacy standards like HIPAA and GDPR, alongside the integration of bias mitigation methods, will be crucial in ensuring fairness and equity in AI-driven healthcare. Moreover, model interpretability will become a priority to foster clinician trust, requiring the development of transparency tools and explainability frameworks. Looking ahead, real-time data integration, especially from wearable devices and IoT sensors, will enable more personalized and predictive healthcare. Furthermore, as the regulatory environment matures, there is a need for global harmonization to ensure that these technologies can be deployed seamlessly across borders. By addressing these challenges, multimodal AI can significantly enhance patient outcomes, streamline healthcare operations, and enable the next frontier in personalized medicine.

Table 6. Applications and Limitations of Multimodal in Medical Imaging and Omics Data

Applications	Limitations
Disease pattern analysis, treatment outcome prediction	Limited diverse imaging data, underperformance in rare or complex cases
Symptom identification, diagnostic workflows, virtual health consultations	Inaccurate answers in complex or rare cases, limited real-time data
Gene-disease relationship insights, research trend identification	Misses recent research, lacks multi-faceted genomic and text analysis
Quick comprehension of patient imaging records, radiology interpretation support	Limited to NLP-based patterns, missing subtle details in imaging data
Precision medicine, personalized treatment plans, advanced diagnostic insights	High computational resources required, potential genomic misinterpretation
Pathology report generation, linking tissue-level data with clinical findings	Poor generalization in diverse tissue types, biased histology data
X-ray and imaging feature analysis, enhanced disease detection capabilities	Overgeneralizes uncommon conditions, difficulty with radiography variations
Multi-imaging disease identification (X-ray, MRI, CT), improved diagnostic accuracy	Limited high-detail scan support, error-prone in detecting subtle features
Complex biomolecular pathway analysis, cross-omics insights for disease research	Scalability issues with real-world omics complexity, oversimplification of complex data
Holistic patient assessments, personalized health predictions	Overconfidence in predictions, limited explainability in genomic-based conclusions

References

1- D. Ramazzotti, A. Lal, B. Wang, S. Batzoglou, and A. Sidow, "Multi-omic tumor data reveal diversity of molecular mechanisms that correlate with survival." *Nat Commun,* Vol. 9 (No. 1), p. 4453, Oct 26 (2018).

2- R. J. Chen *et al.*, "Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis." *IEEE Trans Med Imaging,* Vol. 41 (No. 4), pp. 757-70, Apr (2022).

3- Jie Hao, Sai Chandra Kosaraju, Nelson Zange Tsaku, Dae Hyun Song, and Mingon Kang, "PAGE-Net: Interpretable and Integrative Deep Learning for Survival Analysis Using Histopathological Images and Genomic Data." Presented at the *Biocomputing 2020*, (2019).

4- T. Bhattacharya *et al.*, "AI Meets Exascale Computing: Advancing Cancer Research With Large-Scale High Performance Computing." *Front Oncol,* Vol. 9p. 984, (2019).

5- F. Mohsen, H. Ali, N. El Hajj, and Z. Shah, "Artificial intelligence-based methods for fusion of electronic health records and imaging data." *Sci Rep,* Vol. 12 (No. 1), p. 17981, Oct 26 (2022).

6- Md Rezaul Karim, Tanhim Islam, Christoph Lange, Dietrich Rebholz-Schuhmann, and Stefan Decker, "Adversary-Aware Multimodal Neural Networks for Cancer Susceptibility Prediction From Multiomics Data." *IEEE Access,* Vol. 10pp. 54386-409, (2022).

7- K. Chaudhary, O. B. Poirion, L. Lu, and L. X. Garmire, "Deep Learning-Based Multi-Omics Integration Robustly Predicts Survival in Liver Cancer." *Clin Cancer Res,* Vol. 24 (No. 6), pp. 1248-59, Mar 15 (2018).

8- K. Sone *et al.*, "Application of artificial intelligence in gynecologic malignancies: A review." *J Obstet Gynaecol Res,* Vol. 47 (No. 8), pp. 2577-85, Aug (2021).

9- H. Arimura, M. Soufi, H. Kamezawa, K. Ninomiya, and M. Yamada, "Radiomics with artificial intelligence for precision medicine in radiation therapy." *J Radiat Res,* Vol. 60 (No. 1), pp. 150-57, Jan 1 (2019).

10- M. Efremova and S. A. Teichmann, "Computational methods for single-cell omics across modalities." *Nat Methods,* Vol. 17 (No. 1), pp. 14-17, Jan (2020).

11- H. Xiao, Zhou, F., Liu, X., Liu, T., Li, Z., Liu, X., & Huang, X, "A Comprehensive Survey of Large Language Models and Multimodal Large Language Models in Medicine." DOI: 10.48550/arxiv.2405.08603

12- D. A. Elhag *et al.*, "Inflammatory Bowel Disease Treatments and Predictive Biomarkers of Therapeutic Response." *Int J Mol Sci,* Vol. 23 (No. 13), Jun 23 (2022).

13- Laura Antonelli, Mario Rosario Guarracino, Lucia Maddalena, and Mara Sangiovanni, "Integrating imaging and omics data: A review." *Biomedical Signal Processing and Control,* Vol. 52pp. 264-80, (2019).

14- K. M. Boehm, P. Khosravi, R. Vanguri, J. Gao, and S. P. Shah, "Harnessing multimodal data integration to

advance precision oncology." *Nat Rev Cancer,* Vol. 22 (No. 2), pp. 114-26, Feb (2022).

15- University Radiology Associates. PET/CT Scan. [Online]. Available: https://www.upstate.edu/ura/pet-ct-scan.php.

16- Wikipedia contributors. (2023). H&E stain. [Online]. Available: https://en.wikipedia.org/w/index.php?title=H%26E_stain&oldid=1191484566.

17- Using FAST on Whole-Slide Images (WSI). [Online]. Available: https://fast.eriksmistad.no/python-tutorial-wsi.html.

18- Supporting Multi-omics Approaches. (August 27). [Online]. Available: https://www.thermofisher.com/uk/en/home/brands/thermo-scientific/molecular-biology/molecular-biologylearning-center/molecular-biology-resource-library/spotlight-articles/supporting-multi-omics-approaches.html.

19- C. Massard *et al.*, "High-Throughput Genomics and Clinical Outcome in Hard-to-Treat Advanced Cancers: Results of the MOSCATO 01 Trial." *Cancer Discov,* Vol. 7 (No. 6), pp. 586-95, Jun (2017).

20- W. Silva, & Dos, S. S. W, "Review on multi-modal AI models to integrate imaging and omics data." *Utrecht University*, (2024).

21- National Cancer Institute. (24 Aug). NCI Genomic Data Commons. [Online]. Available: https://gdc.cancer.gov/.

22- The Cancer Imaging Archive (TCIA). National Cancer Institute. [Online]. Available: https://imaging.cancer.gov/informatics/cancer_imaging_archive.htm.

23- R. Sun *et al.*, "A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study." *Lancet Oncol,* Vol. 19 (No. 9), pp. 1180-91, Sep (2018).

24- Sina Tabakhi, Mohammod Naimul Islam Suvon, Pegah Ahadian, and Haiping Lu, "Multimodal Learning for Multi-omics: A Survey." *World Scientific Annual Review of Artificial Intelligence,* Vol. 01(2023).

25- National Genomics Data Center. EWAS Datahub. [Online]. Available: https://ngdc.cncb.ac.cn/ewas/datahub.

26- CPTAC | Office of Cancer Clinical Proteomics Research. National Cancer Institute. [Online]. Available: https://proteomics.cancer.gov/programs/cptac.

27- Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics." *Nat Rev Genet,* Vol. 10 (No. 1), pp. 57-63, Jan (2009).

28- S. C. Sealfon and T. T. Chu, "RNA and DNA microarrays." *Methods Mol Biol,* Vol. 671pp. 3-34, (2011).

29- ENCODE Project Consortium. (August 23). [Online]. Available: https://www.encodeproject.org/.

30- A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and Hjwl Aerts, "Artificial intelligence in radiology." *Nat Rev Cancer,* Vol. 18 (No. 8), pp. 500-10, Aug (2018).

31- UK Biobank. (August 23). [Online]. Available: https://www.ukbiobank.ac.uk/.

32- L. M. Gibson *et al.*, "Factors associated with potentially serious incidental findings and with serious final diagnoses on multi-modal imaging in the UK Biobank Imaging Study: A prospective cohort study." *PLoS One,* Vol. 14 (No. 6), p. e0218267, (2019).

33- A. Partin, Brettin, T., Zhu, Y., Dolezal, J. M., Kochanny, S., Pearson, A. T., Shukla, M., Evrard, Y. A., Doroshow, J. H., & Stevens, R. L., "Data augmentation and multimodal learning for predicting drug response in patient-derived xenografts from gene expressions and histology images." DOI: 10.48550/arXiv.2204.11678

34- Patient-Derived Models Repository (PDMR). [Online]. Available: https://pdmr.cancer.gov/.

35- National Institutes of Health. (August 23). NIH LINCS Program. [Online]. Available: https://lincsproject.org/.

36- P. Bankhead, & others. (August 20). QuPath. [Online]. Available: https://qupath.readthedocs.io/en/stable/

37- J. M. Dolezal, Kochanny, S., Dyer, E., Srisuwananukorn, A., Sacco, M., Howard, F. M., Li, A., Mohan, P., & Pearson, A. and T, "Slideflow: Deep Learning for Digital Histopathology with Real-Time Whole-Slide Visualization." DOI: https://doi.org/10.48550/arXiv.2304.04142

38- E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley, "Color transfer between images." *IEEE Computer Graphics and Applications,* Vol. 21 (No. 4), pp. 34-41, (2001).

39- Robert Tibshirani, "Regression Shrinkage and Selection Via the Lasso." *Journal of the Royal Statistical Society Series B: Statistical Methodology,* Vol. 58 (No. 1), pp. 267-88, (1996).

40- J. Yan, S. L. Risacher, K. Nho, A. J. Saykin, and L. I. Shen, "Identification of Discriminative Imaging Proteomics Associations in Alzheimer's Disease Via a Novel Sparse Correlation Model." *Pac Symp Biocomput,* Vol. 22pp. 94-104, (2017).

41- E. M. Airoldi, E. A. Erosheva, S. E. Fienberg, C. Joutard, T. Love, and S. Shringarpure, "Reconceptualizing the classification of PNAS articles." *Proc Natl Acad Sci U S A,* Vol. 107 (No. 49), pp. 20899-904, Dec 7 (2010).

42- L. Schneider *et al.*, "Integration of deep learning-based image analysis and genomic data in cancer pathology: A systematic review." *Eur J Cancer,* Vol. 160pp. 80-91, Jan (2022).

43- Zabir Al Nazi and Wei Peng, "Large Language Models in Healthcare and Medical Domain: A Review." *Informatics,* Vol. 11 (No. 3), (2024).

44- E. R. Watson, A. Taherian Fard, and J. C. Mar, "Computational Methods for Single-Cell Imaging and Omics Data Integration." *Front Mol Biosci,* Vol. 8p. 768106, (2021).

45- Raja Vavekanand, Kira Sam, Suresh Kumar, and Teerath Kumar, "CardiacNet: A Neural Networks Based Heartbeat Classifications using ECG Signals." *Studies in Medical and Health Sciences,* Vol. 1 (No. 2), pp. 1-17, (2024).

46- X. Jiang, Yan, L., Vavekanand, R., & Hu, M, "Large Language Models in Healthcare Current Development and Future Directions." DOI: 10.20944/preprints202407.0923.v1

47- T. Zhong, Zhao, W., Zhang, Y., Pan, Y., Dong, P., Jiang, Z., Kui, X., Shang, Y., Yang, L., Wei, Y., Yang, L., Chen, H., Zhao, H., Liu, Y., Zhu, N., Li, Y., Wang, Y., Yao, J., Wang, J., . . . Zhang, T. , "ChatRadioValuer: a Chat large language model for generalizable radiology report generation based on multi-institution and multi-system data." DOI: 10.48550/arxiv.2310.05242

48- J. Zhou *et al.*, "Pre-trained multimodal large language model enhances dermatological diagnosis using SkinGPT-4." *Nat Commun,* Vol. 15 (No. 1), p. 5649, Jul 5 (2024).

49- S. Zhou *et al.*, "Large language models for disease diagnosis: a scoping review." *NPJ Artif Intell,* Vol. 1 (No. 1), p. 9, (2025).

50- F. Gaber *et al.*, "Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis." *NPJ Digit Med,* Vol. 8 (No. 1), p. 263, May 9 (2025).

51- M. Mittermaier, M. M. Raza, and J. C. Kvedar, "Bias in AI-based models for medical applications: challenges and mitigation strategies." *NPJ Digit Med,* Vol. 6 (No. 1), p. 113, Jun 14 (2023).

52- S. Carey, A. Pang, and M. Kamps, "Fairness in AI for healthcare." *Future Healthc J,* Vol. 11 (No. 3), p. 100177, Sep (2024).

53- M. A. Ricci Lara, R. Echeveste, and E. Ferrante, "Addressing fairness in artificial intelligence for medical imaging." *Nat Commun,* Vol. 13 (No. 1), p. 4581, Aug 6 (2022).

54- Emilio Ferrara, "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies." *Sci,* Vol. 6 (No. 1), (2023).

55- H. Zeng, L. Chen, Y. Huang, Y. Luo, and X. Ma, "Integrative Models of Histopathological Image Features and Omics Data Predict Survival in Head and Neck Squamous Cell Carcinoma." *Front Cell Dev Biol,* Vol. 8p. 553099, (2020).

56- G. Puliani *et al.*, "NETest: A Systematic Review Focusing on the Prognostic and Predictive Role." *Neuroendocrinology,* Vol. 112 (No. 6), pp. 523-36, (2022).

57- Mutlu Cukurova, Carmel Kent, and Rosemary Luckin, "Artificial intelligence and multimodal data in the service of human decision‐making: A case study in debate tutoring." *British Journal of Educational Technology,* Vol. 50 (No. 6), pp. 3032-46, (2019).

58- Raja Vavekanand, Ganesh Kumar, and Shakhlokhon Kurbanova, "A lightweight physics-conditioned diffusion multi-model for medical image reconstruction." *Biomedical Engineering Communications,* Vol. 5 (No. 2), (2026).

59- M. R. Farlow, A. Hake, J. Messina, R. Hartman, J. Veach, and R. Anand, "Response of patients with Alzheimer disease to rivastigmine treatment is predicted by the rate of disease progression." *Arch Neurol,* Vol. 58 (No. 3), pp. 417-22, Mar (2001).

60- Wenjing Gao, Yuanyuan Pei, Huiying Liang, Junjian Lv, Jiale Chen, and Wei Zhong, "Multimodal AI System for the Rapid Diagnosis and Surgical Prediction of Necrotizing Enterocolitis." *IEEE Access,* Vol. 9pp. 51050-64, (2021).

61- Yexian Zhang *et al.*, "ELMO: An Efficient Logistic Regression-Based Multi-Omic Integrated Analysis Method for Breast Cancer Intrinsic Subtypes." *IEEE Access,* Vol. 8pp. 5121-30, (2020).

62- Shao W Wang T, Huang Z, Tang H, Zhang J, Ding Z, Huang K, "MORONET: Multi-omics Integration via Graph Convolutional Networks for Biomedical Data Classification." DOI: https://doi.org/10.1101/2020.07.02.184705

63- Weiwei Li, Aniruddha Bhattacharjya, and Xin Feng, "Prediction and analysis of attribute reduction algorithm in rough set in Alzheimer's disease." Presented at the *International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023)*, (2023).

64- B. ElKarami, A. Alkhateeb, H. Qattous, L. Alshomali, and B. Shahrrava, "Multi-omics Data Integration Model Based on UMAP Embedding and Convolutional Neural Network." *Cancer Inform,* Vol. 21p. 11769351221124205, (2022).

65- Ayush Garg and Deepika Bansal, "Application of Machine Learning in Disease Prediction." *International Journal for Research in Applied Science and Engineering Technology,* Vol. 11 (No. 2), pp. 47-51, (2023).