



MULTIMODAL GENERATIVE AI FOR PRECISION MEDICINE: INTEGRATING MEDICAL IMAGING, CLINICAL RECORDS AND GENOMIC DATA

Ms. Sanjana Sekhawat, Dr. Parul Vyas

Department of Chemistry, St. John's College, Dr. Bhimrao Ambedkar University, Agra, UP, India

Abstract

Precision medicine aims to move healthcare from generalized treatment strategies toward patient-specific diagnosis, risk assessment, prognosis, and therapeutic selection. However, clinical decision-making depends on heterogeneous information distributed across medical images, electronic health records, laboratory measurements, genomic profiles, pathology reports, medications, and longitudinal clinical histories. Conventional artificial intelligence systems often process these modalities independently, limiting their ability to capture clinically meaningful relationships across different sources of patient information. This paper proposes a multimodal generative artificial intelligence framework that integrates medical imaging, clinical records, and genomic data within a unified precision-medicine architecture. The proposed framework employs modality-specific encoders for imaging, structured and unstructured clinical information, and genomic features, followed by a multimodal fusion mechanism and a generative reasoning layer. A large multimodal model is used to generate patient-level clinical summaries, risk assessments, biomarker interpretations, and candidate treatment insights while preserving links to the underlying evidence. The framework incorporates retrieval-augmented generation, uncertainty estimation, explainability, privacy protection, and human-in-the-loop validation to address the risks associated with generative AI in healthcare. A comprehensive evaluation methodology is proposed using diagnostic discrimination, calibration, sensitivity, specificity, F1-score, survival prediction, treatment-response prediction, explanation fidelity, hallucination rate, fairness, and robustness. The paper discusses challenges including missing modalities, heterogeneous data formats, genomic privacy, dataset shift, demographic bias, hallucination, clinical validation, and regulatory requirements. The proposed framework provides a research foundation for developing trustworthy multimodal generative AI systems capable of synthesizing complementary biomedical evidence to support individualized clinical decision-making.

Keywords: Multimodal Generative AI, Precision Medicine, Medical Imaging, Electronic Health Records, Genomics, Multimodal Learning, Large Language Models, Clinical Decision Support, Personalized Medicine, Healthcare AI

INTRODUCTION

Healthcare generates enormous amounts of heterogeneous data. Modern patients may have radiological images, pathology slides, laboratory results, medication histories, physician notes, genomic sequencing data, vital signs, and longitudinal electronic health records (EHRs). Each modality provides a different perspective on the patient's biological and clinical condition.

Medical imaging provides spatial and anatomical information. Clinical records contain symptoms, diagnoses, medications, procedures, laboratory findings, and longitudinal observations. Genomic information provides insight into molecular characteristics, inherited variation, somatic mutations, and potentially treatment-relevant biomarkers.

Precision medicine seeks to combine these heterogeneous observations to characterize individual patients and support patient-specific healthcare decisions.



Recent research describes multimodal AI as a promising approach for combining genomics, imaging, and EHR data to improve disease detection, risk stratification, biomarker discovery, and clinical decision support.

However, integrating these data types remains challenging. Imaging data are high-dimensional spatial representations, EHRs contain both structured and free-text information, while genomic data are extremely high-dimensional and have specialized biological representations. Consequently, simply concatenating all available features does not necessarily produce clinically meaningful representations.

Generative AI introduces a new possibility. Instead of using AI solely for classification, generative models can synthesize information from multiple modalities and produce clinically understandable outputs.

For example, given:

- a CT or MRI scan,
- a patient's clinical history,
- laboratory findings,
- genomic variants,

a multimodal AI system could potentially produce a structured patient summary and identify relationships among imaging findings, clinical characteristics, and molecular information.

Recent literature has identified the progression from text-based large language models toward multimodal systems capable of processing combinations of imaging, text, and structured medical information.

This paper therefore proposes a multimodal generative AI architecture for precision medicine that integrates three major information sources:

Medical Imaging+Clinical Records+Genomic Data
Medical\ Imaging + Clinical\ Records + Genomic\ Data
The objective is not to replace clinicians but to create an evidence-grounded computational framework for synthesizing complementary patient information.

PROBLEM STATEMENT

Current clinical AI systems frequently focus on individual modalities.

For example:

- imaging AI detects abnormalities in radiological images;
- clinical NLP models extract information from medical notes;
- genomic models identify molecular biomarkers.

Although these specialized models can be useful, independent analysis can overlook interactions between modalities.

Consider a cancer patient with:

1. an imaging finding indicating tumor progression;
2. a clinical history indicating previous treatment;
3. a genomic alteration associated with a particular molecular pathway.

Analyzing these observations independently may fail to capture their combined significance.

The research problem is therefore:

How can medical imaging, longitudinal clinical records, and genomic information be integrated into a multimodal generative AI system capable of producing evidence-grounded and clinically interpretable patient-level insights?



RESEARCH OBJECTIVES

The objectives of this research are:

1. To develop a multimodal architecture integrating medical images, clinical records, and genomic data.
2. To design modality-specific representation learning mechanisms.
3. To develop a multimodal fusion mechanism for patient-level representation.
4. To incorporate a generative AI reasoning layer.
5. To integrate external biomedical knowledge through retrieval-augmented generation.
6. To generate evidence-grounded clinical summaries and risk assessments.
7. To quantify uncertainty and detect unsupported model outputs.
8. To evaluate fairness, robustness, privacy, and clinical reliability.
9. To establish a human-in-the-loop framework for clinical decision support.

RESEARCH CONTRIBUTIONS

The proposed research makes the following contributions:

4.1 Unified Multimodal Architecture

A common architecture is proposed for integrating:

- medical imaging;
- clinical records;
- genomic information.

4.2 Cross-Modal Representation

The framework learns relationships among visual, textual, structured clinical, and molecular information.

4.3 Generative Clinical Reasoning

The model produces patient-level summaries rather than only numerical classifications.

4.4 Evidence-Grounded Generation

Retrieval-augmented generation is used to connect generated outputs to biomedical knowledge and patient evidence.

4.5 Uncertainty-Aware Decision Support

The system explicitly represents uncertainty rather than presenting every prediction as certain.

4.6 Safety-Oriented Design

The framework incorporates privacy protection, bias assessment, hallucination detection, human validation, and clinical governance.

BACKGROUND AND RELATED WORK

Precision Medicine

Precision medicine aims to consider individual differences in genetics, environment, lifestyle, and clinical characteristics when preventing, diagnosing, and treating disease.

Traditional clinical decision-making primarily depends on information available through individual tests and professional interpretation.

AI provides an opportunity to analyze much larger combinations of patient information.



Recent work on multimodal AI in precision medicine emphasizes the integration of genomic, transcriptomic, proteomic, imaging, environmental, and EHR data.

MEDICAL IMAGING IN PRECISION MEDICINE

Medical imaging provides spatial information about biological structures.

Examples include:

- X-ray;
- CT;
- MRI;
- PET;
- ultrasound;
- mammography;
- digital pathology.

An imaging encoder can transform an image I into a latent representation:

$$z_I = f_I(I) \quad z_I = f_I(I)$$

where f_I is an imaging encoder.

Depending on the application, the encoder may use:

- convolutional neural networks;
- vision transformers;
- hierarchical vision transformers;
- foundation vision models.

The representation can capture:

- anatomical structures;
- lesions;
- tissue characteristics;
- morphological abnormalities;
- spatial relationships.

Recent research on multimodal generative medical-image interpretation demonstrates the growing interest in combining image understanding with generative models that can produce clinically meaningful narratives.

CLINICAL RECORDS

EHR data contain both structured and unstructured information.

Structured information

Examples include:

- age;
- sex;
- laboratory values;
- vital signs;
- diagnoses;
- medications;
- procedures;
- hospitalization records.

Unstructured information

Examples include:

- physician notes;
- discharge summaries;



- pathology reports;
- radiology reports;
- consultation notes.

A clinical encoder transforms the available information into:

$$z_C = f_C(C) \quad z_{_C} = f_{_C}(C)$$

where CC represents the clinical record.

Large language models can be used to process clinical text, but their outputs require careful grounding because generated statements may not always be supported by patient records.

GENOMIC DATA

Genomic information provides molecular-level information.

Possible inputs include:

- single-nucleotide variants;
- copy-number variations;
- gene-expression profiles;
- somatic mutations;
- germline variants;
- epigenetic information;
- molecular signatures.

A genomic encoder produces:

$$z_G = f_G(G) \quad z_{_G} = f_{_G}(G)$$

where GG represents genomic data.

Because genomic datasets can contain millions of features, dimensionality reduction or biologically informed feature selection may be required.

Possible approaches include:

- gene-level aggregation;
- pathway-level representation;
- autoencoders;
- transformer-based genomic encoders;
- graph-based molecular representations.

PROPOSED MULTIMODAL ARCHITECTURE

The proposed framework consists of six principal components:

1. Imaging encoder
2. Clinical encoder
3. Genomic encoder
4. Multimodal fusion module
5. Generative reasoning model
6. Evidence and safety layer

The architecture can be represented as:

$$I \rightarrow EI \rightarrow Z_I \quad \text{\textbackslash} \rightarrowtail \quad E_I \quad \text{\textbackslash} \rightarrowtail \quad Z_{_I} \quad C \rightarrow EC \rightarrow Z_{CC} \quad \text{\textbackslash} \rightarrowtail \quad E_{_C} \quad \text{\textbackslash} \rightarrowtail \quad Z_{_C} \\ G \rightarrow EG \rightarrow Z_{GG} \quad \text{\textbackslash} \rightarrowtail \quad E_{_G} \quad \text{\textbackslash} \rightarrowtail \quad Z_{_G}$$

followed by:

$$(Z_I, Z_C, Z_G) \rightarrow \text{Fusion} \rightarrow Z_M \rightarrow \text{Generative Model}(Z_I, Z_{_C}, Z_{_G}) \quad \text{\textbackslash} \rightarrowtail \quad \text{Fusion} \quad \text{\textbackslash} \rightarrowtail \quad Z_{_M} \\ \text{\textbackslash} \rightarrowtail \quad \text{Generative} \backslash \text{Model}$$

where:

- II = medical imaging;
- CC = clinical records;



- GG = genomic data;
- ZI,ZC,ZGZ_I,Z_C,Z_G = modality representations;
- ZMZ_M = multimodal representation.

MULTIMODAL FUSION

Early Fusion

In early fusion, representations are concatenated:

$$ZM=[ZI;ZC;ZG]Z_M=[Z_I;Z_C;Z_G]$$

Although simple, this approach may not adequately model interactions between modalities.

CROSS-ATTENTION FUSION

A more flexible strategy is cross-modal attention.

For two modalities:

$$\text{Attention}(Q,K,V)=\text{Softmax}(QK^T/d)V \quad \text{Attention}(Q,K,V) = \text{Softmax} \left(\frac{QK^T}{\sqrt{d}} \right) V$$

The model can learn relationships such as:

- imaging features $\leftrightarrow$ clinical symptoms;
- imaging features $\leftrightarrow$ genomic alterations;
- genomic features $\leftrightarrow$ treatment history.

A multimodal representation can then be written as:

$$ZM=\text{Fusion}(ZI,ZC,ZG)Z_M=\text{Fusion}(Z_I,Z_C,Z_G)$$

where the fusion module learns cross-modal dependencies.

MISSING-MODALITY HANDLING

Real-world patients rarely have complete multimodal datasets.

One patient may have:

- imaging + EHR;
- another may have EHR + genomics;
- another may have all three.

Therefore, the model should support:

$$ZM=\text{Fusion}(ZI,ZC,ZG,M)Z_M=\text{Fusion}(Z_I,Z_C,Z_G,M)$$

where MM represents a modality-availability mask.

This prevents the system from assuming that every patient has identical data.

Possible approaches include:

- modality dropout during training;
- masked attention;
- modality-specific adapters;
- missing-modality reconstruction;
- uncertainty-aware fusion.

GENERATIVE AI LAYER

The generative model receives the multimodal patient representation:

$$ZMZ_M$$

and produces a clinical output:

$$Y=\text{LLM}(ZM,R)Y=\text{LLM}(Z_M,R)$$

where:



- YY = generated clinical output;
- RR = retrieved biomedical knowledge.

Possible outputs include:

- patient summary;
- risk assessment;
- biomarker interpretation;
- differential considerations;
- treatment-response hypothesis;
- clinical research summary.

The system should distinguish between **evidence**, **model inference**, and **uncertain hypotheses**.

RETRIEVAL-AUGMENTED GENERATION

A major challenge with generative medical AI is hallucination.

To reduce unsupported generation, the proposed framework uses Retrieval-Augmented Generation (RAG).

The retrieval system can access:

- peer-reviewed biomedical literature;
- clinical guidelines;
- drug information;
- genomic databases;
- molecular pathway databases;
- institutional clinical protocols.

The process is:

Patient Data → Query Generation → Knowledge Retrieval → Evidence Filtering → Generative Model
 Patient\ Data \rightarrow Query\ Generation \rightarrow Knowledge\ Retrieval \rightarrow Evidence\ Filtering \rightarrow Generative\ Model

The generated response should contain links between claims and supporting evidence where technically feasible.

PATIENT-LEVEL MULTIMODAL REPRESENTATION

The multimodal representation can be defined as:

$$Z_P = \alpha Z_I + \beta Z_C + \gamma Z_G \quad Z_P = \alpha Z_I + \beta Z_C + \gamma Z_G$$

where:

$$\alpha + \beta + \gamma = 1 \quad \alpha + \beta + \gamma = 1$$

and the coefficients represent learned modality contributions.

However, fixed weights are insufficient for all clinical situations.

Therefore, the proposed framework uses dynamic attention:

$$[\alpha, \beta, \gamma] = \text{Softmax}(W[Z_I, Z_C, Z_G]) \quad [\alpha, \beta, \gamma] = \text{Softmax}(W[Z_I, Z_C, Z_G])$$

This enables the system to adapt to the information available for each patient.

PRECISION RISK PREDICTION

A downstream prediction model can estimate:

$$P(Y|Z_P)P(Y|Z_P)$$

where YY may represent:

- disease risk;
- recurrence risk;
- treatment response;



- survival probability;
- adverse-event risk.

For example:

$$\hat{Y} = \text{MLP}(Z_P) \quad \hat{Y} = \text{MLP}(Z_P)$$

The generative model then converts the numerical output into an interpretable explanation.

PERSONALIZED TREATMENT MODELING

A treatment-response model can be represented as:

$$P(R|T, Z_P)P(R|T, Z_P)$$

where:

- RR = treatment response;
- TT = candidate treatment;
- ZPZ_P = patient representation.

The model could estimate relative response patterns for multiple candidate interventions.

Importantly, such outputs should be considered **decision-support information**, not autonomous prescriptions.

EXPLAINABLE AI

Explainability is essential in precision medicine because clinicians need to understand the evidence supporting an AI-generated recommendation.

The proposed framework provides explanations at three levels.

Level 1: Imaging Explanation

The system identifies regions contributing to a prediction.

Examples include:

- lesion regions;
- tissue areas;
- anatomical structures.

Level 2: Clinical Explanation

The system identifies relevant:

- symptoms;
- laboratory values;
- diagnoses;
- medications;
- clinical events.

Level 3: Genomic Explanation

The system identifies:

- relevant variants;
- genes;
- molecular pathways;
- biomarkers.

The final explanation combines the three.



MULTIMODAL EVIDENCE ATTRIBUTION

For prediction YY, the contribution of each modality can be estimated as:

$$C_m = \frac{\Delta Y_m}{\sum_j |\Delta Y_j|} \quad C_m = \frac{\Delta Y_m}{\sum_j |\Delta Y_j|}$$

where C_m represents the relative contribution of modality mm.

For example:

Modality	Example Evidence
Imaging	Lesion characteristics
Clinical	Laboratory and treatment history
Genomic	Molecular alteration
Combined	Cross-modal evidence

This helps clinicians understand whether a prediction was primarily influenced by imaging, clinical history, genomic information, or their combination.

GENERATIVE CLINICAL REPORT

The proposed model can generate a structured report containing:

Patient Context

Summary of relevant clinical history.

Imaging Findings

Important imaging observations.

Molecular Findings

Relevant genomic characteristics.

Integrated Interpretation

Relationships across the three modalities.

Risk Assessment

Predicted risk and uncertainty.

Evidence

Supporting clinical or biomedical information.

Limitations

Missing or uncertain evidence.

Clinician Review

A section requiring professional verification.

This structure prevents the system from presenting a single unrestricted paragraph as the complete clinical conclusion.



UNCERTAINTY ESTIMATION

A clinical AI system should communicate uncertainty.

Let:

$$U=f(Z_I,Z_C,Z_G)U=f(Z_I,Z_C,Z_G)$$

represent predictive uncertainty.

Uncertainty can increase when:

- modalities are missing;
- input quality is poor;
- the patient differs from the training population;
- conflicting evidence exists;
- genomic variants are uncertain;
- imaging findings are ambiguous.

The system can therefore output:

Prediction+Confidence+Evidence+Uncertainty
Prediction + Confidence + Evidence + Uncertainty
rather than prediction alone.

DATA PRIVACY

Healthcare data contain highly sensitive information.

Genomic information is particularly sensitive because it can contain information relevant to individuals and biological relatives.

The proposed architecture should therefore incorporate:

- encryption;
- access control;
- de-identification;
- secure data storage;
- audit logging;
- data minimization;
- privacy-preserving computation where appropriate.

NIST's AI Risk Management Framework emphasizes managing risks across the AI lifecycle and supports trustworthy and responsible AI development and deployment.

FEDERATED MULTIMODAL LEARNING

When multiple hospitals participate, centralized pooling of patient-level data may not be feasible.

A federated extension can be defined as:

$$\theta_{t+1} = \sum_{k=1}^K \frac{n_k}{N} \theta_{k,t+1} \quad \theta_{k,t+1} = \theta_k + \eta \nabla_{\theta_k} \ell_k(\theta_k)$$

where each institution trains locally.

This allows participating institutions to contribute to model development without necessarily transferring raw patient records.

However, federated learning introduces additional issues such as:

- client heterogeneity;
- model-update privacy;
- poisoning;
- communication costs;
- institutional data differences.



BIAS AND FAIRNESS

A multimodal model can inherit biases from training data.

Potential sources include:

- underrepresentation of populations;
- differences in healthcare access;
- hospital-specific imaging equipment;
- genomic population differences;
- documentation differences;
- socioeconomic disparities.

Fairness should therefore be evaluated across relevant demographic and clinical subgroups.

For subgroup g :

$$TPR_g = \frac{TP_g}{TP_g + FN_g}$$

Differences between groups can be monitored.

However, fairness should not be reduced to one numerical metric; clinical relevance and population characteristics must also be considered.

CLINICAL SAFETY AND REGULATORY CONSIDERATIONS

Multimodal AI intended to support clinical decisions requires careful validation.

The FDA's current digital-health guidance includes a final January 2026 guidance on Clinical Decision Support Software and provides guidance on when software functions may or may not fall within medical-device regulation.

The FDA also emphasizes lifecycle management for AI-enabled medical-device software and notes the importance of managing AI/ML technologies throughout development, deployment, and maintenance.

For this reason, the proposed framework should incorporate:

- predefined intended use;
- patient-population definition;
- data-quality requirements;
- external validation;
- monitoring after deployment;
- version control;
- model-change management;
- clinician oversight.

PROPOSED EXPERIMENTAL METHODOLOGY

Because this manuscript proposes a research framework, actual numerical performance should be obtained through implementation rather than fabricated.

26.1 Dataset Design

A suitable experimental dataset should contain, where possible:

$$D = \{I, C, G, Y\}$$

where:

- I = imaging;
- C = clinical information;
- G = genomic information;
- Y = clinical outcome.

Potential application areas include:



- oncology;
- cardiovascular disease;
- neurological disorders;
- rare diseases.

Cancer is particularly suitable because imaging, clinical history, pathology, and genomic information frequently interact in diagnosis and treatment research.

Recent reviews report increasing application of imaging-genomics fusion across organs and clinical areas, while highlighting representation, cross-dataset integration, and multimodal data challenges.

EXPERIMENTAL PIPELINE

The experimental pipeline should include:

Phase 1 — Data Preparation

- de-identification;
- quality control;
- normalization;
- missing-value handling.

Phase 2 — Modality Encoding

- imaging encoder;
- clinical encoder;
- genomic encoder.

Phase 3 — Multimodal Fusion

Cross-attention or another fusion architecture.

Phase 4 — Prediction

Disease/risk/treatment-response prediction.

Phase 5 — Generation

Generate patient-level evidence-based summaries.

Phase 6 — Validation

Evaluate clinical and technical performance.

BASELINE MODELS

The proposed model can be compared with:

1. Imaging-only deep learning
2. Clinical-only model
3. Genomics-only model
4. Imaging + clinical model
5. Imaging + genomics model
6. Clinical + genomics model
7. Conventional multimodal neural network
8. Multimodal transformer
9. Multimodal generative model
10. Proposed multimodal generative AI framework

This comparison determines whether combining all three modalities provides additional predictive information.

EVALUATION METRICS

29.1 Classification

Accuracy:



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall:

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1-score:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

29.2 AUROC

Area Under the Receiver Operating Characteristic Curve can be used for discrimination.

29.3 AUPRC

AUPRC is particularly informative for imbalanced clinical outcomes.

29.4 Calibration

Calibration should assess whether predicted probabilities correspond to observed outcome frequencies.

Possible measures include:

- Brier score;
- calibration slope;
- calibration intercept.

GENERATIVE AI EVALUATION

Generative outputs require additional evaluation.

Factuality

Are generated statements supported by patient data?

Groundedness

Are claims supported by retrieved biomedical evidence?

Hallucination Rate

$$\text{HR} = \frac{\text{Unsupported Claims}}{\text{Total Claims}}$$

Clinical Completeness

Does the generated report include important findings?

Consistency

Does the model provide consistent conclusions when equivalent inputs are supplied?

Expert Evaluation

Clinicians can rate:

- relevance;
- accuracy;
- usefulness;
- clarity;
- safety.

ABLATION STUDY

The following ablation configurations can be evaluated:

Model Imaging Clinical Genomics Generative AI

M1	✓	–	–	–
M2	–	✓	–	–
M3	–	–	✓	–



	Model	Imaging	Clinical	Genomics	Generative AI
M4	✓	✓	–	–	
M5	✓	–	✓	–	
M6	–	✓	✓	–	
M7	✓	✓	✓	–	
M8	✓	✓	✓	✓	

M8 represents the complete proposed architecture.

The purpose is to determine the contribution of each modality and the generative reasoning layer.

ROBUSTNESS EVALUATION

The model should be tested under:

- missing imaging;
- missing genomic data;
- incomplete EHRs;
- low-quality images;
- noisy clinical records;
- uncertain genomic variants;
- demographic shifts;
- hospital-to-hospital distribution shifts.

The goal is to determine whether the model remains reliable when real-world data differ from ideal research datasets.

EXPECTED RESEARCH OUTCOMES

The proposed research hypothesizes that multimodal integration can provide a more comprehensive representation of patients than single-modality models.

Potential advantages include:

Improved Patient Representation

The model can combine biological, clinical, and anatomical information.

Improved Risk Stratification

Cross-modal relationships may provide information unavailable to individual modalities.

Better Evidence Synthesis

Generative AI can organize heterogeneous findings into structured reports.

Clinical Workflow Support

Automated synthesis may reduce the time required to review large volumes of patient information.

Biomarker Discovery

Relationships between imaging phenotypes and molecular features may support biomarker research.

These are hypotheses requiring empirical validation rather than established outcomes of the proposed system.



CHALLENGES

34.1 Data Heterogeneity

Images, clinical notes, and genomic data have fundamentally different structures.

34.2 Missing Data

Not every patient has all modalities.

34.3 Data Alignment

The imaging date, genomic test date, and clinical record may not correspond exactly.

34.4 Genomic Privacy

Genomic data require particularly careful privacy and governance.

34.5 Hallucination

Generative models may produce plausible but unsupported clinical statements.

34.6 Dataset Bias

Models can perform differently across populations and healthcare institutions.

34.7 Interpretability

A multimodal model may be difficult to explain because evidence is distributed across multiple modalities.

34.8 Clinical Validation

High benchmark performance does not automatically demonstrate clinical utility.

FUTURE RESEARCH DIRECTIONS

35.1 Multimodal Medical Foundation Models

Large pretrained models could provide general-purpose representations across medical modalities.

35.2 Digital Patient Twins

Multimodal AI could potentially support computational representations of patient trajectories.

35.3 Federated Multimodal Learning

Multiple hospitals could collaboratively train models while retaining local patient data.

35.4 Longitudinal Patient Modeling

Future systems should integrate information across months or years rather than analyzing isolated encounters.

35.5 Causal AI

Causal approaches could help distinguish associations from clinically meaningful causal relationships.

35.6 Multimodal Drug Response Prediction

Imaging, genomic, and clinical information could be investigated jointly for treatment-response modeling.

35.7 Human-AI Collaboration

Future studies should evaluate how clinicians interact with multimodal AI explanations and whether these systems improve decision-making without increasing cognitive or automation bias.

DISCUSSION

The central concept of this research is that precision medicine requires **integration rather than isolated prediction**.

Medical imaging describes the patient's anatomical and phenotypic characteristics.

Clinical records describe:

- symptoms;
- disease history;
- treatment;
- laboratory findings;
- clinical trajectory.

Genomic data provide molecular information.

The proposed framework combines these sources:

$$\text{Patient} = \text{Imaging} + \text{Clinical} + \text{Genomic} \quad \text{Patient} = \text{Imaging} + \text{Clinical} + \text{Genomic}$$

and transforms them into a unified representation:

$$ZP = F(ZI, ZC, ZG) \quad Z_P = F(Z_I, Z_C, Z_G)$$

The generative AI component then transforms the representation into clinically interpretable information.

However, generative capability should not be confused with clinical authority. A generated statement is not automatically medically valid merely because it is fluent or plausible.

Therefore, the proposed architecture emphasizes:

$$\text{Evidence} \rightarrow \text{Prediction} \rightarrow \text{Explanation} \rightarrow \text{Human Validation} \quad \text{Evidence} \rightarrow \text{Prediction} \rightarrow \text{Explanation} \rightarrow \text{Human Validation}$$

rather than:

$$\text{AI} \rightarrow \text{Automatic Clinical Decision} \quad \text{AI} \rightarrow \text{Automatic Clinical Decision}$$

This distinction is particularly important for high-stakes healthcare applications.

Current FDA guidance recognizes that software functions analyzing medical images or providing patient-specific clinical decision support may have regulatory implications depending on their intended use and functionality.

CONCLUSION

This paper proposed a multimodal generative artificial intelligence framework for precision medicine that integrates medical imaging, clinical records, and genomic information. The proposed architecture addresses a major limitation of conventional healthcare AI: the tendency to analyze different sources of patient information independently.

The framework uses modality-specific encoders to transform imaging, clinical, and genomic information into representations that can be integrated through cross-modal attention. A generative AI layer then synthesizes the multimodal representation into patient-level clinical summaries, risk assessments, biomarker interpretations, and evidence-grounded decision-support information.

Retrieval-augmented generation is incorporated to improve evidence grounding, while uncertainty estimation, explainability, privacy protection, fairness assessment, and human oversight address important safety requirements.

The proposed approach has potential applications in oncology, cardiovascular medicine, neurology, rare diseases, and other areas in which clinical phenotype, longitudinal records, and molecular information jointly influence patient management. Recent research supports the increasing importance of multimodal



integration across imaging, EHRs, and genomics, while also identifying interoperability, bias, privacy, and validation as major challenges.

Future experimental studies should evaluate the framework using appropriately matched multimodal datasets and independent external cohorts. Performance should be assessed not only through conventional predictive metrics but also through calibration, robustness, fairness, hallucination rate, evidence grounding, explanation quality, and clinician evaluation.

Ultimately, multimodal generative AI should be developed as a **clinician-support technology**, with transparent evidence, explicit uncertainty, appropriate validation, and controlled deployment. Such an approach can help realize the potential of AI-driven precision medicine while reducing the risks associated with opaque or unsupported automated clinical reasoning.

REFERENCES

1. Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1. National Institute of Standards and Technology. <https://doi.org/10.6028/NIST.AI.100-1>.
2. Shahper Nazeer Khan, S. N., Danishuddin, Mohd Wajid Ali Khan, et al. (2026). Multi-modal AI in precision medicine: integrating genomics, imaging, and EHR data for clinical insights. *Frontiers in Artificial Intelligence*, 8, 1743921. <https://doi.org/10.3389/frai.2025.1743921>.
3. Wang, S., Liu, M., Li, Y., Zhang, X., Sun, M., Wang, Z., Li, R., Li, Q., He, Y., Hu, X., Sun, L., Yan, F., Yu, M., Ding, W., & Wang, C. (2025). Recent advances in data-driven fusion of multimodal imaging and genomics for precision medicine. *Information Fusion*, 115, 102738. <https://doi.org/10.1016/j.inffus.2024.102738>.
4. *Harnessing Artificial Intelligence in Multimodal Omics Data Integration: Paving the Path for the Next Frontier in Precision Medicine*. (2024). *Annual Review of Biomedical Data Science*.
5. Ghebrehiwet, I., Zaki, N., Damseh, R., et al. (2024). Revolutionizing personalized medicine with generative AI: a systematic review. *Artificial Intelligence Review*, 57, 128. <https://doi.org/10.1007/s10462-024-10768-5>.
6. Rao, V. M., Hla, M., Moor, M., Adithan, S., Kwak, S., Topol, E. J., Rajpurkar, P., et al. (2025). Multimodal generative AI for medical image interpretation. *Nature*, 639, 888–896.
7. *From large language models to multimodal AI: a scoping review on the potential of generative AI in medicine*. (2025). *Biomedical Engineering Letters*.
8. U.S. Food and Drug Administration. (2026). *Clinical Decision Support Software: Guidance for Industry and Food and Drug Administration Staff*. FDA.
9. U.S. Food and Drug Administration. (2026). *Guidances with Digital Health Content*. FDA.
10. U.S. Food and Drug Administration. (2025). *Artificial Intelligence in Software as a Medical Device*. FDA.
11. U.S. Food and Drug Administration. (2022). *Guidance on the Application of ISO 14971 to Artificial Intelligence and Machine Learning*. FDA Recognized Consensus Standards.
12. U.S. Food and Drug Administration. *Clinical Decision Support Software: Digital Health Policy Navigator*. FDA.
13. NIST. (2023). *AI Risk Management Framework Playbook*. National Institute of Standards and Technology.