Multimodal Foundation Models and Quantum-Accelerated AI for Drug Discovery and Precision Healthcare: Survey and Framework

Maria Michael Visuwasam L
Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia.
ponraj.me@gmail.com
Dept of ADS, RMK College of Eng.amd Technology,
mariamichaelads@rmkcet.ac.in

Pawan Kumar Chaurasia
Babasaheb Bhimrao Ambedkar Central University Lucknow Uttar Pradesh India.
pkc.gkp@gmail.com

Abstract
The rapid expansion of heterogeneous biomedical data—including molecular graphs, protein conformations, genomic sequences, medical imaging, and clinical records—has intensified the demand for artificial intelligence (AI) frameworks capable of unified and biologically grounded analysis. Although recent deep learning advances have achieved notable success within individual biomedical domains, most existing approaches remain unimodal or weakly multimodal, resulting in fragmented representations and limited generalization across complex biological systems. In drug discovery, generative models such as variational autoencoders, generative adversarial networks, and diffusion-based approaches show promise in molecular design but often lack explicit protein structure conditioning, limiting biochemical validity and target specificity. Graph neural networks support molecular representation learning but face scalability challenges due to increasing architectural complexity. Quantum machine learning has been proposed to enhance molecular property prediction, yet its application remains largely theoretical. In clinical settings, predictive models predominantly rely on correlational learning, reducing interpretability and cross-institutional robustness. This paper presents a comprehensive survey and unified problem formulation highlighting the need for multimodal biomedical foundation models integrating generative chemistry, quantum–classical hybrid learning, and causal inference to advance next-generation drug discovery and precision healthcare.

Keywords: Multimodal Foundation Models; Drug Discovery; Generative Models; Diffusion Models; Graph Neural Networks; Quantum Machine Learning; Causal Inference; Precision Healthcare.

Introduction

The biomedical and healthcare domains are undergoing a profound transformation driven by the rapid growth of heterogeneous data generated from diverse sources, including molecular structures, protein conformations, genomic and transcriptomic profiles, medical imaging, and electronic clinical records. Advances in high-throughput sequencing, structural biology, medical imaging technologies, and digital health systems have resulted in unprecedented data volumes and complexity. Effectively integrating and reasoning over this heterogeneous information is critical for accelerating drug discovery, improving therapeutic efficacy, and enabling precision healthcare.

Artificial intelligence (AI) and machine learning have emerged as powerful tools for extracting patterns from biomedical data and supporting decision-making across drug discovery and clinical applications. Deep learning models have achieved remarkable success in tasks such as molecular property prediction, protein structure estimation, disease

diagnosis from medical images, and clinical outcome forecasting. Despite these advances, the majority of existing approaches operate within narrowly defined problem settings and are designed for single data modalities. Such unimodal or weakly multimodal systems are inherently limited in their ability to capture the complex, multiscale interactions that characterize biological systems, where molecular, structural, genomic, and clinical factors jointly influence therapeutic outcomes.

Recent progress in foundation models has demonstrated the feasibility of learning general-purpose representations from large-scale data that can be adapted to multiple downstream tasks. While foundation models have transformed fields such as natural language processing and computer vision, their adoption in biomedicine remains fragmented. Existing biomedical models often lack unified architectures capable of jointly modeling chemical structures, protein interactions, and clinical phenotypes. Moreover, many multimodal learning approaches rely on shallow fusion strategies, such as feature concatenation, which inadequately capture cross-modal dependencies and consequently limit generalization and robustness.

In drug discovery, generative modeling techniques based on variational autoencoders, generative adversarial networks, and diffusion processes have been proposed to explore vast chemical spaces and design novel compounds. Although these approaches have demonstrated the ability to generate chemically valid molecules, many lack explicit conditioning on protein structural information, thereby restricting target specificity and biological relevance. Graph neural networks have further advanced molecular representation learning by modeling relational dependencies within chemical graphs; however, their increasing computational complexity and reliance on classical optimization frameworks pose scalability challenges when applied to large molecular libraries.

Parallel to these developments, quantum machine learning has emerged as a promising paradigm for enhancing molecular representation and accelerating property estimation through quantum–classical hybrid computation. Despite theoretical advantages, practical adoption remains limited, with most studies confined to small benchmark datasets and isolated experimental settings. Similarly, clinical AI systems predominantly rely on correlational learning, resulting in limited interpretability, reduced robustness, and poor cross-institutional generalization due to the absence of causal reasoning mechanisms.

Taken together, these observations reveal that existing AI methodologies address individual components of the biomedical pipeline in isolation, without a coherent framework capable of integrating heterogeneous data modalities, biologically grounded generative modeling, computationally efficient learning paradigms, and robust clinical inference. The lack of such integration motivates the need for a formal and unified problem formulation that precisely characterizes the limitations of current approaches and delineates the requirements of next-generation biomedical AI systems.

Organization of the Paper

The remainder of this paper is organized as follows. Section II presents a formal problem statement, precisely defining the key challenges associated with multimodal biomedical data integration, generative molecular modeling, scalable learning, and clinical inference. Section III provides a comprehensive review of related work, systematically surveying recent advances in multimodal foundation models, generative approaches for drug discovery, graph neural networks for molecular representation learning, quantum machine learning

techniques, and causal inference methods in healthcare AI. Section IV discusses the synthesized research gaps and open challenges identified from the literature. Finally, Section V concludes the paper and outlines directions for future research.

II. Problem Statement

Let $\mathcal{D} = \{\mathcal{D}_m, \mathcal{D}_p, \mathcal{D}_g, \mathcal{D}_i, \}$

represent heterogeneous biomedical data modalities, where D_m denotes molecular structures, $D_p$protein_conformations, D_g genomic profiles, D_i medical imaging data, and D_c clinical records. Each modality differs substantially in structure, dimensionality, noise characteristics, and semantic representation. The overarching objective of artificial intelligence–driven drug discovery and precision healthcare is to learn unified and transferable representations from D that support biologically meaningful molecular generation, accurate molecular property prediction, and robust clinical outcome inference.

Despite rapid advances in biomedical machine learning, existing systems exhibit several fundamental limitations. First, most approaches remain unimodal or weakly multimodal, resulting in fragmented representations that fail to capture complex cross-modal dependencies inherent to biological systems. Recent multimodal biomedical models demonstrate progress, yet they lack the scale, alignment mechanisms, and robustness required to function as general-purpose foundation models [1]–[3].

Second, generative models for drug discovery—such as variational autoencoders, generative adversarial networks, and diffusion-based frameworks—have shown promise in exploring chemical space; however, many lack explicit conditioning on protein structural information, thereby limiting target specificity and biochemical plausibility of generated molecules [4]–[7]. Third, graph neural networks have become central to molecular representation learning, but their increasing architectural complexity and reliance on classical optimization introduce scalability and computational constraints when applied to large molecular libraries [8], [9].

Fourth, quantum machine learning has been proposed as a potential accelerator for molecular embedding and property estimation through quantum–classical hybrid computation; nevertheless, current studies remain largely theoretical or restricted to small benchmark datasets, with limited integration into end-to-end biomedical pipelines [10], [11]. Finally, clinical prediction models predominantly rely on correlational learning, leading to reduced interpretability, limited robustness, and poor generalization across institutions due to the absence of causal reasoning [12]–[14].

Accordingly, the central problem addressed in this work is the lack of a unified, scalable, and interpretable framework capable of integrating multimodal biomedical data, enabling target-conditioned molecular generation, supporting computationally efficient learning, and facilitating robust clinical inference within a coherent foundation model paradigm.

Section III: Literature Review

Table 1.0 Compartive Analysis of Previous Works

Category	Key Techniques	Strengths	Limitations	Research Gap Identified	Role in Proposed Framework

Category	Key Techniques	Strengths	Limitations	Research Gap Identified	Role in Proposed Framework
Multimodal Foundation Models	Transformers, BiomedCLIP, multimodal pretraining	Unified representation learning across text, images, and clinical data; improved downstream diagnostics	Limited modality coverage; shallow fusion; weak molecular–protein integration; domain imbalance issues	Lack of end-to-end biomedical foundation models integrating molecular + clinical modalities	Backbone for unified multimodal biomedical intelligence
Generative Models for Drug Discovery	VAE, GAN, Diffusion Models	Effective exploration of chemical space; improved molecule validity; diffusion models offer training stability	Weak protein-structure conditioning; limited biological relevance; poor target specificity	Missing integration with structure-aware target-guided design pipelines	Enables target-aware molecular generation
Graph Neural Networks (GNNs)	MPNNs, hybrid GNNs, multitask learning architectures	Strong molecular topology learning; state-of-the-art property prediction performance	Scalability issues; computational complexity; benchmarking inconsistencies	Inefficient scaling to ultra-large chemical datasets; reliance on classical optimization	Provides molecular embedding layer for structural learning
Quantum Machine Learning (QML)	Variational quantum circuits, quantum kernels, hybrid quantum–classical models	Potential acceleration in molecular simulation and energy prediction	Hardware noise; small dataset validation; lack of real-world deployment pipelines	Limited integration into multimodal biomedical AI systems	Enhances optimization and molecular property prediction
Causal Inference in Clinical AI	Structural causal models, counterfactual reasoning, treatment effect estimation	Improves interpretability, robustness, and bias correction	Mostly independent from foundation models; limited scalability	Absence of causal reasoning within multimodal biomedical architectures	Enables explainable and transferable clinical predictions

IV. Research Gaps

Despite significant progress in multimodal learning, generative modeling, graph neural networks, quantum machine learning, and causal inference, several critical challenges remain in biomedical AI. Current foundation models cannot effectively integrate molecular, protein, genomic, imaging, and clinical data within a unified architecture, leading to fragmented representations and weak cross-modal understanding. Generative drug discovery models often lack protein-structure conditioning, reducing biological relevance and target specificity. Graph neural networks and multimodal architectures face scalability limitations due to computational complexity and large dataset requirements. Quantum machine learning shows theoretical advantages but remains constrained by hardware limitations and limited real-world validation. Clinical AI systems largely rely on correlational learning, resulting in reduced interpretability and weak cross-institutional generalization. Furthermore, causal reasoning is rarely integrated with large-scale foundation models. Existing approaches therefore operate mostly in isolation rather than within unified frameworks. Addressing these limitations requires scalable multimodal architectures with biologically grounded generative capabilities. Integrating quantum–classical learning and causal inference into foundation models is essential for robust biomedical intelligence. Such unified frameworks are critical for advancing next-generation drug discovery and precision healthcare systems.

V. Conclusion and Future Research Directions

This paper presented a structured literature survey and formal problem formulation for artificial intelligence–driven drug discovery and precision healthcare. The review examined recent developments in multimodal foundation models, generative molecular modeling, graph neural networks, quantum machine learning, and causal inference, and demonstrated that existing approaches largely evolve in isolation. As a result, current biomedical AI systems remain limited in their ability to jointly model heterogeneous biomedical data, generate biologically relevant drug candidates, scale efficiently, and provide robust clinical inference.

The analysis identified several critical limitations: (i) the absence of unified multimodal foundation models spanning molecular to clinical modalities; (ii) insufficient incorporation of protein or target conditioning in generative drug discovery frameworks; (iii) scalability and efficiency constraints in graph-based and large multimodal models; (iv) limited practical integration and benchmarking of quantum machine learning techniques; and (v) the predominant reliance on correlational learning in clinical AI systems, resulting in limited interpretability and generalization.

Future research should prioritize the development of scalable multimodal pretraining strategies that support robust cross-modal alignment and transferability. Advancing target-conditioned generative models with explicit protein structure integration is essential for improving biological relevance in drug discovery. In parallel, rigorous evaluation of hybrid quantum–classical learning approaches within end-to-end biomedical pipelines is required to assess their practical advantages. Finally, embedding causal inference mechanisms into large-scale multimodal models represents a key direction for achieving interpretable and generalizable clinical decision support.

References

[1] S. Zhang, Y. Zhang, J. Ma, X. Li, and J. Sun, “BiomedCLIP: A multimodal biomedical foundation model pretrained from fifteen million scientific image–text pairs,” *arXiv preprint arXiv:2303.00915*, 2023.

[2] S. C. Huang, M. Pareek, R. Seyyedi, I. Banerjee, and M. P. Lungren, “Multimodal foundation models for healthcare: A systematic review and implementation guidelines,” *NPJ Digital Medicine*, vol. 8, no. 1, pp. 1–15, 2025.

[3] K. Sun, Y. Liu, H. Wang, and L. Chen, “Medical multimodal foundation models in clinical diagnosis: Progress and challenges,” *Artificial Intelligence in Medicine*, vol. 148, p. 102694, 2025.

[4] A. Alakhdar, M. Atz, and G. Schneider, “Diffusion models in de novo drug design,” *Journal of Chemical Information and Modeling*, vol. 64, no. 2, pp. 405–417, 2024.

[5] U. Das, R. Jain, and A. Gupta, “Generative artificial intelligence for drug discovery and protein design,” *Trends in Pharmacological Sciences*, vol. 46, no. 1, pp. 23–37, 2025.

[6] J. Yim and B. Park, “Diffusion models for protein structure modeling and molecular docking,” *WIREs Computational Molecular Science*, vol. 14, no. 2, p. e1702, 2024.

[7] H. Wang, L. Zhang, and Z. Xu, “ProtDiff: Function-conditioned masked diffusion models for protein design,” in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2024, pp. 1–14.

[8] H. Fooladi, M. Dehghani, and A. Emami, “Evaluating machine learning models for molecular property prediction: Challenges and best practices,” *Journal of Chemical Information and Modeling*, vol. 65, no. 1, pp. 112–126, 2025.

[9] M. Bin Javaid, A. Ullah, and S. Khan, “Multi-task graph neural networks for molecular property prediction,” *Computers & Chemical Engineering*, vol. 182, p. 108493, 2025.

[10] W. Li, J. Romero, and A. Aspuru-Guzik, “Hybrid quantum–classical machine learning of molecular energies and properties,” *Digital Discovery*, vol. 4, no. 1, pp. 55–67, 2025.

[11] J. Romero and A. Aspuru-Guzik, “Toward improved quantum machine learning for molecular simulations,” *arXiv preprint arXiv:2505.04122*, 2025.

[12] L. Jiao, Y. Chen, and J. Pearl, “Causal inference meets deep learning: A comprehensive review,” *Science Partner Journal: Patterns*, vol. 5, no. 3, p. 100612, 2024.