Review

DOI: 10.5582/ddt.2026.01042

Advances in multimodal data integration for drug efficacy prediction: Methodological evolution and clinical translation

Jingwen Fang[1], Zihan Wang[1], Junhao Shao[3], Jiewei Fan[1], Qichao Luo[2,*]

[1] School of Health Management, Anhui Medical University, Hefei, Anhui, China;
[2] School of Pharmacy, Anhui Medical University, Hefei, Anhui, China;
[3] The Second Affiliated Hospital of Anhui Medical University, Hefei, Anhui, China.

SUMMARY: Drug efficacy prediction remains a cornerstone of drug development and precision therapy. However, integrating heterogeneous biomedical data, including multi-omics profiles, pathological imaging, electronic health records, and pharmacokinetic-pharmacodynamic (PK/PD) time-series, faces three fundamental barriers, namely cross-domain distribution shifts between preclinical and clinical data, relational mismatches between isolated vector representations and biological networks, and feature heterogeneity across disparate modalities. To address these challenges, three AI paradigms have emerged, transfer learning for cross-domain alignment, graph neural networks for structured relational modeling, and Transformers for global cross-modal feature interaction. Importantly, these techniques form a many-to-many complementary system rather than a one-to-one correspondence, a key insight that this review explicitly formalizes. We further elaborate encoding workflows for PK/PD data to bridge static molecular signatures with dynamic *in vivo* exposure trajectories. Four graded clinical applications are outlined, including personalized monotherapy, combination optimization, drug repurposing, and preclinical-to-clinical evaluation of novel candidates. We also dissect persistent bottlenecks such as data harmonization, model interpretability, and prospective validation, and propose five actionable directions, namely privacy-preserving benchmarks, causally interpretable models, temporal dynamic frameworks, cross-domain generalization, and lightweight clinical tools. By integrating theoretical rationales, methodological synergies, and hierarchical translational scenarios, this review provides a unified roadmap to accelerate the clinical deployment of multimodal drug response prediction.

Keywords: Multimodal fusion, drug efficacy prediction, transfer learning, graph neural network, transformer, PK/PD modeling, clinical translation

1. Introduction

Drug efficacy prediction critically influences drug development, clinical decision-making, and precision medicine. Conventional pipelines relying on *in vitro* assays, animal models, and small-scale trials are time-consuming, costly, and prone to high attrition. In clinical practice, "one-size-fits-all" regimens often fail owing to profound inter-patient genetic, phenotypic, and physiological heterogeneity, leading to suboptimal responses, adverse events, and accelerated resistance (*1-3*). Computational biology and artificial intelligence (AI) have offered powerful alternatives. Early models primarily used single-omics data, including genomics, transcriptomics, proteomics, and metabolomics, to decipher molecular mechanisms, enabling pharmacogenomic-guided dosing (*4-6*). However, biological systems are dynamic, multi-scale networks spanning molecules to organisms. Drug responses are modulated by gene expression, the tumor microenvironment, tissue architecture, and systemic factors, information that single-omics approaches cannot fully capture. This fragmented view largely explains the poor translational success of preclinical findings (*7,8*).

Multimodal data integration has thus emerged as a transformative paradigm, unifying multi-omics profiles, pathology images, electronic health records (EHRs), molecular structures, and pharmacokinetic/pharmacodynamic (PK/PD) time-series. Yet such integration confronts three intertwined bottlenecks, namely the cross-domain gap, systematic distribution shifts between *in vitro* cell lines and clinical patients; the relational gap, the mismatch between isolated feature vectors and the networked nature of biological interactions; and the feature gap, the inherent heterogeneity in semantics, scale, and distribution across modalities (9). Three mainstream AI modeling paradigms have been widely adopted to resolve these fusion

challenges, with each addressing one core statistical gap as its primary advantage while simultaneously mitigating multiple auxiliary limitations. Transfer learning mainly alleviates cross-domain covariate shift and limited labeled sample sizes. Graph neural networks (GNNs) provide natural structural frameworks to quantify networked biological correlations among genes, proteins, small-molecule drugs, and disease phenotypes. Transformer architectures leverage self-attention mechanisms to capture global feature interactions and long-range biological dependencies, strengthening cross-modal information aggregation. Collectively, the three methodologies primarily target three core gaps in multimodal drug response modeling, namely the cross-domain gap, the relational gap, and the feature gap, and form a mutually complementary technical system with overlapping functional boundaries (*10-12*).

This review establishes a complete logical chain covering multimodal heterogeneous data integration, optimized statistical modeling *via* AI algorithms, and real-world clinical translational deployment. Centered on transfer learning, graph neural networks, and Transformer architectures, we systematically summarize the methodological evolution, integrated fusion frameworks, hierarchical clinical application scenarios, and unresolved bottlenecks of multimodal drug efficacy prediction. The manuscript elaborates on three core integration barriers, three mainstream technical pathways, four layered clinical translational scenarios, and three categories of systematic practical challenges. Despite remarkable progress in algorithmic design, prominent obstacles remain unresolved, including the standardized harmonization of heterogeneous multimodal datasets, model interpretability, and the scarcity of prospective clinical validation cohorts, all of which demand further targeted research. This review aims to construct a systematic, forward-looking overview for researchers and clinical practitioners, and to accelerate the translational transformation of computationally predicted drug efficacy from theoretical modeling to routine clinical practice.

2. Literature search and study selection

This manuscript is a narrative review rather than a systematic or scoping review; accordingly, we did not follow the standardized PRISMA-ScR reporting framework. To ensure full transparency in literature collection and selection, we detail the retrieval strategies and screening procedures below. Literature retrieval was performed across four authoritative biomedical databases, namely PubMed, Web of Science Core Collection, CNKI, and Embase, with the search time window set from January 2018 to April 2026 to capture cutting-edge advances in multimodal fusion and AI-driven drug efficacy prediction over the past decade. Landmark foundational papers published before 2018 were manually supplemented by tracing the citation chains of highly cited works. Three groups of combinatorial English keywords were used for retrieval, covering drug efficacy prediction (drug response prediction, drug efficacy forecasting, personalized pharmacotherapy, drug repurposing, and combination drug synergy), multimodal data integration (multi-omics fusion, pathological imaging, electronic health records, molecular structures, and multimodal learning), and artificial intelligence algorithms (transfer learning, graph neural networks, Transformers, covariate shift, domain adaptation, and biological graph modeling). Corresponding Chinese keywords were adopted for CNKI retrieval to incorporate domestic pharmaceutical and biostatistical research.

Clear inclusion and exclusion criteria were established for literature filtering. Eligible studies included original research articles, reviews, and methodological papers concentrating on multimodal fusion modeling for drug response prediction; studies that constructed or validated transfer learning, graph neural network, or Transformer architectures using pharmaceutical datasets; and research involving clinical translational analysis, PK/PD multimodal integration, or cross-domain distribution calibration of drug prediction models. Excluded works comprised pure single-omics studies without multimodal fusion design; papers merely introducing basic AI theories without pharmaceutical or clinical validation; conference abstracts, case reports, editorials, letters, and other non-peer-reviewed grey literature; as well as duplicate publications and articles with incomplete experimental data or ambiguous model descriptions. The screening workflow consisted of two rounds of assessment: titles and abstracts were first scanned to remove irrelevant records, then all remaining full texts were independently evaluated by two authors against the inclusion criteria, with any disagreements resolved through group discussion. Ultimately, 84 representative studies were selected to build the literature foundation of this narrative review.

3. Three core gaps in multimodal data integration

Multimodal data provide comprehensive multi-scale information for drug efficacy prediction, spanning from molecular to phenotypic levels and from microscopic to macroscopic perspectives. However, inherent differences in the sources, structures, and semantics of various data types mean that simple concatenation or shallow fusion fails to achieve effective integration and may even introduce noise and bias. Three fundamental bottlenecks underlie this challenge, as illustrated in Figure 1. The first major barrier is the cross-domain gap, manifested as systematic distribution differences among datasets from distinct biological backgrounds, experimental conditions, and patient populations. The most typical example is the irreducible discrepancy between *in vitro* cell line data and clinical patient data. During long-term

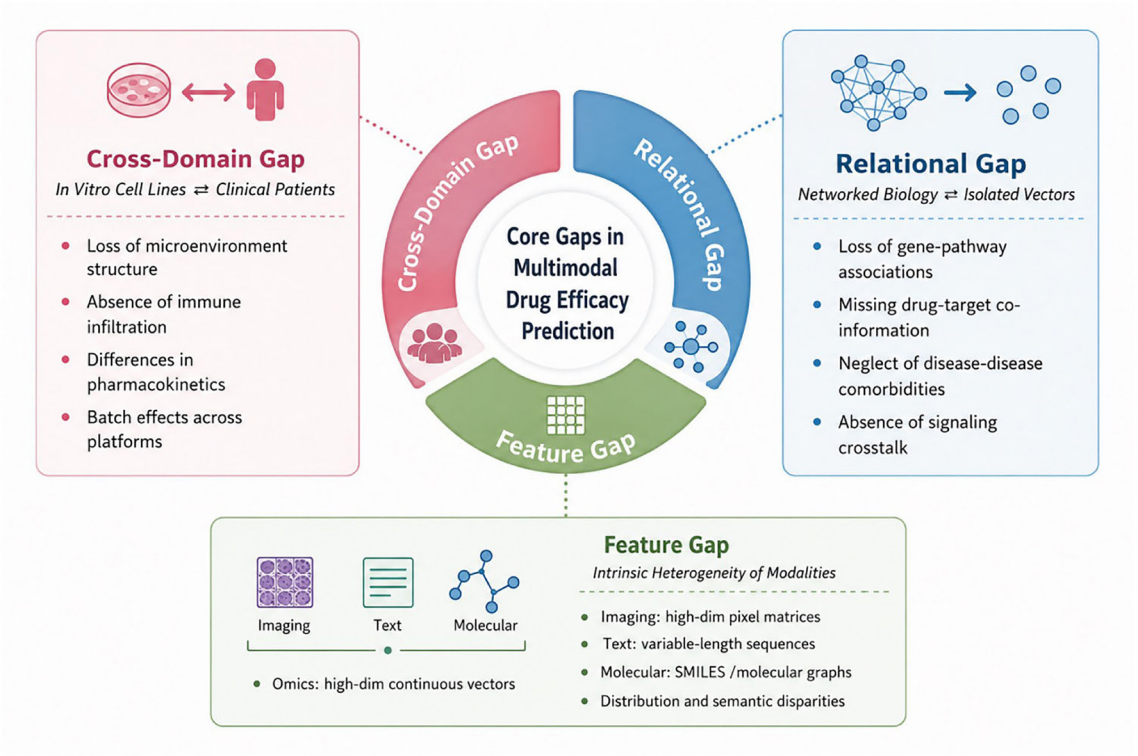

Figure 1. Core gaps in multimodal drug efficacy prediction. The three main challenges include the cross-domain gap between in vitro cell line data and clinical patients, the relational gap between networked biological systems and isolated vector representations, and the feature gap caused by intrinsic heterogeneity across different data modalities (omics, imaging, molecular structures, and text).

passage, cell lines lose the microenvironmental structure, immune infiltration, stromal interaction, and pharmacokinetic characteristics of the original tumor, leading to systematic deviations in their gene expression profiles and drug response spectra compared with patients' primary tumors. Models trained on cell line databases such as the Cancer Cell Line Encyclopedia (CCLE) and the Cancer Therapeutics Response Portal (CTRP) often show significant performance decline when applied to patient-derived xenograft (PDX) models or clinical cohorts (*12,13*). This gap is not merely a statistical distribution difference but also stems from incomplete biological information: cell line data lacks key dimensions such as microenvironment and immune status, which cannot be compensated for by simple data alignment. Similar cross-domain issues also exist between different sequencing platforms, laboratory batches, and population cohorts, severely limiting the generalization ability of predictive models.

The second key obstacle is the relational gap, arising from the mismatch between the network structure of biological systems and conventional isolated data representation. Life activities are not linear superpositions of isolated entities but dynamic networks coupled by multi-level relationships such as gene regulation, protein-protein interactions, and signaling pathway crosstalk (*14*). Drug responses often involve coordinated expression of multiple genes and feedback regulation of pathways. However, traditional data representation methods treat each sample as an independent vector, completely losing association information between entities, such as whether two genes are located in the same pathway, whether drugs share targets, or whether there are comorbidity relationships between diseases. These are precisely the key prior knowledge constraints for predicting drug efficacy. The essence of the relational gap is the structural contradiction between isolated data representation and the networked characteristics of biological systems, which prevents models from using relational priors to constrain the prediction space.

Beyond cross-domain and relational barriers, multimodal integration is further constrained by the feature gap, which arises from inherent heterogeneity across data modalities in terms of physical meaning, dimensional scale, statistical distribution, and semantic level. For instance, drug molecular structures can be encoded as SMILES sequences, molecular fingerprints, or molecular graphs; pathological images are presented as high-dimensional pixel matrices; and clinical data exist as structured electronic health records or free-text narratives. These different modalities vary significantly in their structural forms, dimensional scales, and statistical distributions, making direct feature alignment and fusion difficult (*15-17*). Traditional shallow fusion methods, such as simple feature concatenation or weighted averaging, often fail to exploit the complementary advantages of heterogeneous data and may even introduce noise or dilute key biological signals,

thereby limiting improvement in prediction accuracy.

These three gaps are not isolated but closely coupled and mutually interactive. The cross-domain gap exacerbates distribution discrepancies across different data modalities; the lack of relational information undermines the stability of cross-domain alignment; and semantic barriers at the feature level further hinder the effective integration of multi-source data. Accordingly, systematic advances in multimodal drug efficacy prediction require simultaneous resolution of all three core challenges, along with the construction of an integrated framework that accommodates multi-source data, characterizes biological interactions, and unifies heterogeneous feature representations.

4. Core technical methods for multimodal drug efficacy prediction

Facing the cross-domain gap, relational gap, and feature gap inherent in multimodal data integration, transfer learning, GNNs, and Transformer models provide targeted technical support from different dimensions. Together they form a complete methodological system that realizes horizontal integration of heterogeneous data, structured modeling of biological relationships, and in-depth interaction of multimodal features. Transfer learning mainly addresses the cross-domain discrepancy and bridges data from cell lines and clinical patients *via* knowledge migration. GNNs focus on the relational gap, constructing structured biological frameworks through heterogeneous graphs and hypergraphs. The Transformer architecture targets the feature gap, unifying multimodal representation and cross-modal alignment by leveraging global self-attention. Multimodal fusion relies on the above three techniques to achieve deep feature aggregation and semantic consistency. These methods develop hierarchically and complement one another, driving drug efficacy prediction from single-data analysis toward comprehensive multi-dimensional information fusion. Representative technical methods for addressing these core challenges are summarized in Table 1, and their many-to-many relationships with the three integration gaps are further visualized in Table 2.

4.1. Transfer learning: horizontal knowledge transfer across data sources

Transfer learning serves as the primary paradigm to alleviate the cross-domain gap by constructing knowledge migration pipelines that link *in vitro* cell line source domains with clinical patient target cohorts. Its methodological evolution follows two threads: from homogeneous single-modality migration to cross-domain generalization across distinct distributions, and from bulk-level alignment to cross-scale adaptation at single-cell resolution (*18-20*). The central challenge remains

Table 1. Representative technical methods for addressing core challenges in multimodal drug efficacy prediction

Technical Method	Addressed Core Gap	Core Idea	Key Technical Points
Transfer Learning	Cross-domain gap (cell line → clinical patient distribution shift)	Achieve cross-domain data alignment through knowledge transfer mechanisms	Adversarial domain adaptation; multi-task learning; triplet loss; progressive transfer; cross-resolution integration; contrastive domain regularization
Graph Neural Network	Relational gap (biological network *vs.* isolated feature representation)	Construct structured relational frameworks to model complex associations	Heterogeneous GNN; hypergraph learning; cross-layer graph fusion; graph attention; graph contrastive learning; graph autoencoder-based representation disentanglement.
Transformer	Feature gap (cross-modal heterogeneity and semantic disparity)	Enable cross-modal alignment and long-range dependency modeling *via* self-attention.	Self-attention mechanism; pre-trained language models; multi-head attention; GNN-Transformer fusion; equivariant geometric GNN + Transformer co-modeling.

Table 2. Many-to-many correspondence between mainstream modeling methods and three core integration gaps

Method Category	Cross-Domain Gap	Relational Gap	Feature Gap
Transfer Learning	● Primary	○ Auxiliary	○ Auxiliary
GNN	○ Auxiliary	● Primary	○ Auxiliary
Transformer Architectures	○ Auxiliary	○ Auxiliary	● Primary
Contrastive Learning	○ Auxiliary	○ Auxiliary	● Primary
Adversarial Domain Adaptation	● Primary	—	○ Auxiliary
Graph Autoencoders	—	● Primary	○ Auxiliary
Multi-Task Learning	○ Auxiliary	○ Auxiliary	● Primary
Meta-Learning	● Primary	○ Auxiliary	—

Symbol definitions: ● Primary: The technique is fundamentally developed to address this gap and provides dominant, core resolution strategies. ○ Auxiliary: The method yields secondary, indirect benefits to relieve the gap as an additional functionality. — No direct contribution: The technique exhibits no explicit mitigating effect on the gap, and published evidence supporting such functionality is insufficient at present.

how to reliably extrapolate drug response knowledge from cultured cell lines to clinical populations.

For preclinical translation, ScreenDL leverages patient-derived organoids (PDOs) as intermediate carriers and designs a three-stage progressive workflow covering cell lines, PDX models, and clinical samples (*21*). Pharmaformer validates this strategy using retrospective colorectal and bladder cancer datasets, confirming that intermediate models reduce cross-domain bias (*22*). DTLCDR integrates chemical descriptors, molecular graphs, drug-target interaction profiles, and single-cell transcriptomics to boost out-of-distribution performance (*23*). TransCDR extracts unified latent drug representations *via* cross-domain transfer and achieves stable generalization toward unseen scaffolds (*24*). Clinically, transfer learning repurposes large-scale cell-line screening data for personalized therapy, offering advantages for rare malignancies. MODAE jointly optimizes autoencoder reconstruction, domain adaptation, response regression, and survival prediction (*25*). In renal cell carcinoma studies, deep transfer learning unifies single-cell, spatial transcriptomics, and metabolomics to identify tyrosine kinase inhibitor-tolerant subpopulations (*26*). scXDR enables cross-dataset harmonization *via* graph message passing and domain-invariant alignment (*27*). However, cell lines cannot fully recapitulate the tumor microenvironment, setting an upper bound on accuracy (*28*). Severe domain divergence may cause negative transfer, yet most methods align only marginal distributions without preserving causal structures (*29*). Quantitative metrics for detecting and mitigating negative transfer remain underdeveloped (*30*).

4.2. Graph neural networks: Structured modeling of biological relationships

GNNs construct topological frameworks of multi-type biological entities and accomplish vertical multi-scale aggregation from molecular profiles to clinical phenotypes. Their evolution follows three directions: from homogeneous to heterogeneous graphs, from pairwise to high-order interaction mining, and from flat to hierarchical architectures. Heterogeneous GNNs dominate drug response modeling. MTGNN unifies compounds, targets, and disease phenotypes within one graph and leverages direction-sensitive meta-paths to trace drug-target-disease associations (*11*). MRDDA expands entity sets to include drugs, diseases, proteins, genes, and pathways, combining mixed convolution layers for local and global feature extraction (*31*). GraphTCDR maintains stable fusion performance under partial node feature missing (*32*). Heterogeneous pipelines have also been extended to traditional Chinese medicine research (*33*). Hypergraph neural networks capture collective synergistic effects beyond pairwise graphs. MRHGNN encodes high-order drug-target-protein complex interactions (*34*). HCLGT-DRP constructs modality-specific hypergraphs followed by cross-hypergraph attention aggregation (*35*). Many works incorporate molecular pre-trained models such as ChemBERTa for node feature initialization (*36*). Heterogeneous graphs rely on meta-path reasoning, whereas hypergraphs naturally describe multi-molecule group interactions.

Cross-layer fusion strategies align with biological hierarchy. DeepKGI builds multi-tier gene graphs covering molecule, pathway, and functional module layers, using cross-layer attention to capture cross-scale links (*37*). GLASynergy constructs global-local dual views, extracting local topological signatures *via* GCNs and generating holistic drug-pair representations *via* adaptive attention (*38*). Clinically, GNNs integrate mutation profiles, drug-drug interactions, and signaling crosstalk to guide combinatorial regimen design (*39*). Graph contrastive learning and adaptive attention refine feature screening: AGRL-DSE enables weighted multi-scale aggregation (*40*); RSGCL-DTI mitigates bias on imbalanced datasets (*41*); MOLGAECL combines autoencoder pre-training with contrastive fine-tuning (*42*). Nevertheless, heterogeneous and hypergraph GNNs face scalability bottlenecks, as computational cost grows rapidly with node types and hyperedge orders, limiting deployment on large patient cohorts. Designing lightweight architectures that retain relational expressiveness remains a key engineering challenge for clinical translation.

4.3. Transformer: global alignment of multimodal feature interaction

Transformers mitigate the feature gap *via* self-attention, overcoming the sequential limitations of RNNs and CNNs while delivering a global receptive field. Unlike GNNs that encode explicit topologies, Transformers excel at holistic context modeling and cross-modal semantic alignment. Clinically, Transformers unify multi-omics, pathology images, and EHRs to construct complete patient phenotypic profiles. DualPG-DTA uses pre-trained language encoders to process SMILES strings and protein sequences, with cross-attention enabling bidirectional interaction for binding affinity prediction (*43*). MTEGDRP integrates Transformer with equivariant GNNs: the Transformer captures long-range omics dependencies while equivariant GNNs retain 3D conformational geometry (*44*). For high-order association mining, HCLGT-DRP combines multi-head attention with hypergraphs in a unified Graph Transformer (*35*). Beyond prediction, Transformers improve interpretability; visualized attention maps pinpoint binding residues and functional fragments, guiding clinical decisions (*43*). Hybrid GNN-Transformer architectures preserve topological constraints while harnessing global context, forming a standardized pipeline: the GNN branch aggregates relational

information, the Transformer branch encodes cross-modal dependencies, and bidirectional cross-attention fuses both streams (*44*). Nevertheless, pure Transformers cannot fully encode biomolecular topologies, and stacked attention layers incur high computational costs; attention pruning and graph-Transformer co-design are thus critical for deployment efficiency.

4.4. Synergistic mechanisms and fusion paradigms

These three methodologies form complementary rather than rigid one-to-one relationships, and their organic integration is essential for thorough multimodal fusion. Three fusion paradigms differ in execution order and aggregation logic. Sequential fusion executes modules stepwise: transfer learning first aligns cell-line knowledge, GNNs then build heterogeneous graphs, and Transformers finally perform cross-modal interaction (*10,45*). This offers clear division but risks error propagation. Drug-target affinity prediction typically follows this design (*43,45-47*). Parallel fusion operates modules concurrently within a unified architecture, with attention gating dynamically fusing multi-source representations (*44*). This maximizes complementarity but imposes stricter convergence constraints, and has become mainstream for synergy prediction (*20,34,35*). Hierarchical fusion maps biological layers onto tiered networks; bottom transfer learning harmonizes cross-domain data, middle GNNs preserve relational priors, and top Transformers enable global semantic communication (*20,35*). This best matches the hierarchical nature of biological regulatory systems and represents a leading research direction (*38,48,49*).

Auxiliary algorithms, including contrastive learning, adversarial domain adaptation, graph autoencoders, multi-task learning, and meta-learning, are embedded into these paradigms to resolve sub-problems, each alleviating multiple gaps simultaneously. Table 2 visualizes the many-to-many mapping between all modeling methods and the three core gaps. Nearly all current frameworks assume temporal consistency of multimodal data, yet real-world clinical data exhibit temporal heterogeneity: EHRs are irregular sparse time series, imaging spans multiple visits, and omics are mostly baseline-only. Dynamic GNNs and time-aware Transformers represent promising directions to model longitudinal trajectories and response evolution.

4.5. Integration of pharmacokinetic and pharmacodynamic multimodal data

Pharmacokinetic (PK) and pharmacodynamic (PD) readouts are indispensable multimodal clinical covariates bridging preclinical modeling and individualized outcomes. PK profiles quantify ADME behaviors, while PD signatures capture dose-dependent target inhibition and phenotypic responses. Combined PK-PD datasets contain continuous concentration-time curves, longitudinal dosing records, and graded adverse event biomarkers, which cannot be fully characterized by static omics or molecular graphs alone. Their integration mitigates the feature gap and partially alleviates cross-domain shift from incomplete *in vivo* signals in cell line data (*50,51*). Time-aware Neural-ODE encoders map concentration trajectories into compact embeddings, accommodating irregular sampling (*52-55*). Dose-response regression extracts descriptors such as maximum efficacy and half-maximal concentration, appended as node attributes in GNNs (*56,57*). Categorical treatment records are processed *via* multi-task learning with auxiliary toxicity and efficacy branches (*58,59*). Toxicity endpoints are encoded as ordinal features or integrated into multi-objective losses (*60*). This multi-layer encoding unifies PK-PD temporal signals with static molecular features within shared latent manifolds (*61*).

The three fusion paradigms can incorporate PK/PD branches with differentiated optimization. In sequential fusion, transfer learning harmonizes preclinical omics, GNNs model drug-target-PD associations, and Transformers integrate longitudinal PK trajectories with static signatures (*62*). In parallel fusion, independent encoding branches for molecular graphs, pathology, multi-omics, and PK/PD time-series are dynamically weighted *via* cross-attention, as in MMPK for oral PK prediction (*63*). In hierarchical fusion, the bottom layer calibrates cross-species PK *via* transfer learning, the middle layer encodes drug-pathway relationships *via* GNNs, and the top layer integrates dynamic trajectories *via* Transformers (*64,65*). Empirical validations confirm that PK/PD-integrated models reduce bias, particularly for compounds with high inter-individual variability, and distinguish true synergy from additive effects (*66,67*). Three bottlenecks persist: animal PK parameters exhibit covariate shift relative to human profiles (*50,68*); few graph frameworks explicitly link dosage, PK exposure, and PD phenotypes *via* hyperedges (*69,70*); and PK time-series, PD curves, and static vectors possess vastly different distributions, with simple concatenation diluting temporal signals (*71*). Sparse clinical sampling further introduces noise. Future work should combine dynamic GNNs with time-equivariant Transformers to jointly encode static networks and time-varying trajectories (*72,73*).

5. Clinical translation scenarios of multimodal drug efficacy prediction

The clinical translational value of multimodal artificial intelligence in drug efficacy prediction can be interpreted from two progressive dimensions: the increasing complexity of clinical decision-making and the full life cycle of drug research and development (R&D). It covers four hierarchical application scenarios: selecting

optimal single-drug regimens for individual patients, optimizing combined therapeutic strategies, discovering new indications for approved drugs, and conducting early clinical evaluation of novel drug candidates. Collectively, these scenarios form a closed-loop ecosystem spanning new drug development, personalized medication, and drug repurposing. This section systematically elaborates the above four scenarios, focusing on the enabling logic, implementation pathways, and remaining translational bottlenecks of multimodal drug efficacy prediction technologies. The core clinical application scenarios are illustrated in Figure 2.

5.1. Personalized monotherapy selection

Personalized monotherapy selection aims to identify the optimal single-drug regimen for individual patients. Multimodal models integrate multi-omics, imaging, and clinical data to match patient profiles with drug properties, outperforming population-based approaches and reducing reliance on expensive molecular assays. Chu *et al.* developed the MDD-CoD framework, which adopts a three-stage cascaded decision workflow that unifies clinical phenotypic readouts and multi-attribute drug descriptors to output individualized medication suggestions, surpassing classical baseline models across test cohorts of chronic kidney disease, rheumatoid arthritis, and colorectal carcinoma (*49*). For clear cell renal cell carcinoma, integrated pipelines incorporating germline and somatic genomic variants, histopathological phenotypes, and longitudinal follow-up records have been established to construct recurrence risk scoring systems, supporting joint modeling of molecular biomarkers, tissue morphological traits, and long-term clinical endpoints (*74,75*). The MMPCC architecture combines hematoxylin and eosin (H&E) whole-slide imaging, somatic mutation profiles, and routine clinical records to predict patient-specific drug responses, representing a viable alternative for resource-limited settings where comprehensive genetic testing is not universally accessible (*76,77*).

Despite encouraging retrospective results, personalized prediction frameworks face substantial translational barriers. Most models are trained on single-center, demographically homogeneous retrospective cohorts, resulting in limited out-of-distribution generalization across multi-regional populations and disparate hospital systems, a challenge that directly mirrors the cross-domain gap defined in Section 2. Current pipelines predominantly focus on static pre-treatment risk stratification and rarely incorporate dynamic longitudinal biomarkers, limiting their capacity to support real-time adaptive dosing consistent with PK/PD temporal modeling (Section 4.5). Moreover, most reported performances are validated solely on retrospective datasets, whereas rigorous prospective clinical trial evidence remains scarce. The paucity of high-quality randomized controlled trial (RCT) datasets constitutes the primary bottleneck to large-scale clinical adoption.

5.2. Combination therapy optimization

When monotherapy fails due to tumor heterogeneity or acquired resistance, clinical decision-making shifts to multi-drug regimen optimization (*78-80*). GNNs and hypergraph architectures capture high-order synergies among compounds, targets, and pathways, advancing

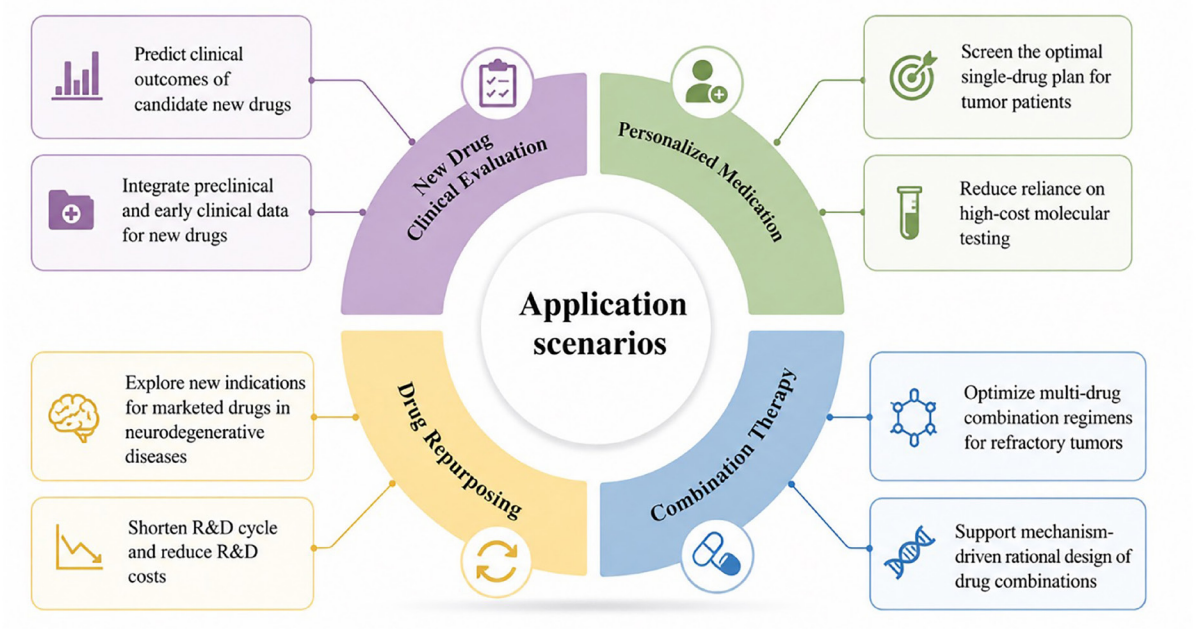

Figure 2. Clinical application scenarios of multimodal drug efficacy prediction. The four core scenarios include personalized medication, combination therapy, drug repurposing, and new drug clinical evaluation, each addressing distinct clinical and translational needs.

combination design from empirical screening to mechanism-guided rational development (*11,45,46*). Huang *et al.* constructed the Madrigal system, which extracts unified latent embeddings from molecular scaffolds, signaling cascade maps, cellular functional readouts, and transcriptomic profiles to quantify combinatorial effects. Validated against a dataset comprising 953 clinical endpoints and 21,842 small molecules, the framework simultaneously predicts adverse pharmacokinetic interactions and maintains concordance with head-to-head clinical trial observations, supplying actionable evidence for combined interventions in type 2 diabetes and metabolic dysfunction-associated steatohepatitis (MASH) (*81*). For synergy regression, Saeed *et al*. proposed MMDGNN, which fuses molecular fingerprint descriptors and SMILES representations within an adaptive graph learning backbone (*80*). Zhang *et al*. developed MOSAIC, a multi-granularity cross-modal encoding framework that accounts for global molecular topology and local substructure fragments, with bidirectional cross-attention gates facilitating information exchange across heterogeneous feature spaces (*81*). Chen *et al.* introduced MRHGNN, a dual-branch hypergraph architecture that mines compound physicochemical attributes and multi-agent high-order interaction topology (*34*). To address data sparsity in rare malignancies, Feng *et al.* proposed MetaSynergy, which implements fast cross-layer meta-adaptive optimization by integrating molecular graph topologies and cell-line multi-omics signatures, effectively mitigating performance degradation arising from limited *in vitro* screening data (*82*).

Despite methodological advances, prominent translational bottlenecks remain. Most computational synergy outputs lack systematic validation in high-fidelity preclinical models, and prospective cohort evidence remains insufficient to establish a complete pipeline from algorithmic prediction to experimental validation and clinical confirmation. Many models output scalar synergy scores without interpretable mechanistic decomposition of molecular interactions, limiting clinical adoption. Current research also displays a strong oncological bias, with far fewer investigations addressing chronic inflammatory and autoimmune disorders. Importantly, only a small proportion of published algorithms have undergone the full translational chain from *in silico* forecasting and laboratory validation to prospective patient verification. High-throughput preclinical platforms based on patient-derived organoids (PDOs) and humanized mouse models have yet to form iterative co-optimization workflows with multimodal prediction architectures. These limitations, spanning insufficient mechanistic interpretability, incomplete disease pathway coverage, and inadequate validation infrastructure, collectively reflect the relational, feature, and cross-domain gaps elaborated in Section 3.

5.3. Drug repurposing

Drug repurposing substantially shortens R&D timelines and reduces costs by identifying new indications for approved agents. GNNs decode drug–disease–target links, multi-omics dissects regulatory cascades, and transfer learning extrapolates knowledge from data-rich to rare diseases, transforming random screening into targeted repurposing (*83*). Gunther *et al.* constructed a multi-stage biologically informed workflow combining transcriptomic profiling, proteomic measurement, and PDO phenotypic functional assays for endometriosis, a chronic disorder with limited non-hormonal options. The framework confirmed elevated ROR1 expression, computationally shortlisted candidates, and completed *in vitro* verification in organoid systems, identifying the marketed agent rimegepant as a promising repurposed candidate (*84*). Beyond wet-lab validation, computational architectures have been deployed across diverse diseases. In neurodegenerative disorders, interpretable deep learning systems incorporating genome-wide association study (GWAS) genotypes and multi-omics signatures prioritize repurposing leads and elucidate molecular pathological cascades (*85,86*). For metabolic and hepatic diseases, multi-omics workflows have uncovered anti-fibrotic bioactivity of existing agents, with compounds validated to modulate the EGR1/STAT3 transcriptional axis (*87*). During viral emergencies, rapid integration of viral genomic signatures and drug-target interaction mechanisms supports high-throughput identification of broad-spectrum repurposed antivirals (*86,88*). The DeepSeq2Drug framework demonstrates the performance gains achieved through multimodal embedding and cross-domain transfer learning for antiviral repurposing (*89*). AI pipelines built on aging hallmark biomarkers enable batch identification of multi-functional agents for chronic disorders and senescence, while also providing a paradigm for discovering clinical indications of herbal constituents from traditional Chinese medicine (TCM) (*90,91*). Despite translational promise, existing repurposing frameworks exhibit structural limitations. Most drug-disease association pipelines rely on phenotypic resemblance or transcriptomic correlation without robust causal molecular evidence. Screened candidates rarely undergo large-sample real-world cohort analysis to confirm long-term efficacy and safety. Current architectures tend to overemphasize therapeutic potency while neglecting clinically critical PK/PD covariates such as optimal dosing, treatment duration, and combinative compatibility. Most critically, many preclinically validated leads fail in late-stage evaluation due to suboptimal pharmacokinetic profiles, off-target toxicities, or inadequate dose-response adaptability. This high attrition rate indicates that the positive predictive value of existing multimodal systems requires substantial enhancement to meet rigorous clinical translational standards.

5.4. Clinical evaluation of novel drug candidates

The core bottleneck in drug development is the predictive discrepancy between preclinical results and clinical outcomes. Multimodal fusion integrates multi-scale preclinical readouts and early-phase clinical data to prospectively forecast efficacy, toxicity, and dose–response profiles before full-scale human trials are initiated (*92,93*). For preclinical efficacy prediction, multi-scale frameworks unify compound topological structures, cellular sensitivity signatures, organoid phenotypic readouts, and *in vivo* PK parameters to build hierarchical modeling architectures spanning molecular, cellular, tissue, and organismal layers. Causal GNNs serve as interpretable backbones to disentangle directional regulatory pathways driving therapeutic responses. The ScreenDL pipeline adopts a staged transfer learning workflow that chains cell line screening, PDX and PDO intermediate platforms, and human clinical cohorts, systematically mitigating prediction bias from interspecies covariate shift (*21*). For early safety and dose optimization, digital twin simulation combined with cross-modal fusion enables preemptive identification of adverse event signals in early trial design, providing quantitative evidence for first-in-human dose escalation strategies (*94*). Beyond efficacy and toxicity forecasting, integrated pipelines incorporating target molecular attributes, historical R&D records of analogous scaffolds, and disease subtype signatures leverage causal reasoning modules to stratify predicted response probabilities across patient subpopulations (*95*).

Three interlinked constraints restrict the maturity of this field. First, stringent proprietary data access barriers and fragmented cross-pharma data standards impede large-scale multi-institutional collaborative dataset construction (*69*). Second, the absence of uniform global benchmark specifications, with divergent normalization rules for molecular encoding, toxicological endpoint quantification, and clinical outcome readouts, hampers cross-dataset transferability and robust out-of-distribution generalization. Third, emerging preclinical models including organoids and organs-on-chips lack systematically curated standardized multimodal training cohorts, restricting iterative model refinement. To address these bottlenecks, industry-academia data-sharing consortia should be established, with privacy-preserving federated learning architectures deployed to support collaborative modeling without raw data exchange. Concurrent development of unified global preclinical-clinical benchmark datasets is an essential prerequisite for scaling the clinical deployment of multimodal novel drug evaluation systems.

6. Core challenges and future breakthrough directions

Despite transformative performance, systematic obstacles persist across data, modeling, and deployment layers, including data heterogeneity and privacy conflicts, model opacity and poor generalization, and incompatibility with clinical workflows. The field lacks unified benchmark standards, interpretability methods remain post-hoc rather than causally embedded, static paradigms cannot capture disease dynamics (*96*), and regulatory frameworks for AI medical tools are still nascent. To address these layered bottlenecks, systematic advances are required across five interconnected dimensions.

Data dimension. Standardized, privacy-preserving multi-center data consortia should be constructed, with federated learning and differential privacy deployed to enable collaborative modeling without raw patient data exchange. Continuous curation of benchmark cohorts derived from organoid and organ-on-a-chip preclinical systems is essential to fundamentally mitigate multimodal heterogeneity and data scarcity limitations (*92,98*).

Model dimension. Causal inference frameworks should be fused with domain-specific biological priors to design natively interpretable graph-based architectures. Simultaneously, biomedical foundation models trained *via* large-scale self-supervised pre-training must be developed to boost cross-distribution generalization capacity (*97,99,100*).

Dynamic temporal dimension. Longitudinal time-series multimodal readouts should be fully exploited to build integrated dynamic GNN-transfer learning pipelines. Such architectures facilitate a paradigm shift from static snapshot efficacy prediction to continuous simulation of progressive disease trajectories and time-variant drug response kinetics (*96*).

Generalization dimension. Large-scale pre-training schemes and cross-domain knowledge transfer modules require further refinement to strengthen model adaptability across variable patient populations, disease spectrums, and real-world clinical settings.

Clinical translational dimension. Lightweight, clinician-friendly analytical modules tailored to hospital IT infrastructure need to be engineered. Closed-loop iterative validation workflows linking preclinical laboratory research and prospective clinical trials ought to be established, alongside gradual refinement of regulatory frameworks governing medical AI devices (*85*).

Beyond these five core directions, deeper integration of cutting-edge architectures including graph Transformers, dynamic graph neural networks, geometric GNNs, and self-supervised graph learning will further expand the methodological boundary of multimodal drug efficacy prediction. Iterative optimization of this integrated technical ecosystem tightens the linkage between early-stage pharmaceutical R&D and individualized clinical treatment. Ultimately, this system can furnish more robust and efficient therapeutic decision schemes for complex malignancies, rare disorders, and emergent public health crises, catalyzing the paradigm

transition from universal "one-size-fits-all" empirical medication to patient-specific precision therapy, and lay solid methodological foundations for fundamental and translational biomedical innovation.

7. Conclusion

This review establishes a unified theoretical framework centered on three core integration bottlenecks, the cross-domain gap, the relational gap, and the feature gap, and comprehensively surveys three mainstream modeling paradigms, transfer learning, graph neural networks, and Transformers, along with their synergistic fusion strategies. A dedicated section on pharmacokinetic and pharmacodynamic (PK/PD) multimodal integration extends the methodological scope to include *in vivo* longitudinal signals, an aspect often underrepresented in previous reviews, while four hierarchical translational scenarios, namely personalized monotherapy, combination optimization, drug repurposing, and novel drug clinical evaluation, form a complete research chain from preclinical modeling to clinical decision-making. Critically, we clarify the many-to-many complementary relationships between modeling algorithms and the three core gaps, moving beyond the oversimplified one-to-one mapping logic adopted in previous summaries, and elaborate three canonical fusion paradigms, sequential, parallel, and hierarchical, with embedded PK/PD time-series trajectory modeling to bridge static multi-omics data and dynamic *in vivo* therapeutic signals. Five targeted breakthrough directions are proposed, including privacy-preserving multi-center benchmark datasets, causally interpretable graph-Transformer architectures, dynamic temporal modeling frameworks, cross-domain generalization strategies, and lightweight clinical tools, with deeper integration of graph Transformers, dynamic hypergraphs, and self-supervised foundation models poised to further expand predictive capabilities. Collectively, this integrated theoretical framework, comprehensive methodological system, and hierarchical clinical application scenarios offer standardized guidance and actionable routes for future multimodal drug response modeling and translational research, with future work prioritizing organ-on-chip combined PK/PD cohorts, causally interpretable dynamic fusion models, and multi-center prospective clinical validation to accelerate the shift from empirical universal medication to individualized precision therapy.

Funding: This work was supported by the Youth Science Fund of the National Natural Science Foundation of China (No.82304382) and University Students' Innovation and Entrepreneurship Training Program (to Junhao Shao).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Yang X, Kui L, Tang M, Li D, Wei K, Chen W. High-throughput transcriptome profiling in drug and biomarker discovery. Front Genet. 2020; 11:19.
2. Han R, Yoon H, Kim G, Lee H, Lee Y. Revolutionizing medicinal chemistry: the application of artificial intelligence (AI) in early drug discovery. Pharmaceuticals (Basel). 2023; 16:1259.
3. Zahra MA, Al-Taher A, Alquhaidan M, Hussain T, Ismail I, Raya I. The synergy of artificial intelligence and personalized medicine for the enhanced diagnosis, treatment, and prevention of disease. Drug Metab Pers Ther. 2024; 39:47-58.
4. Li X, Wang Z, Wang Q, Xu Q, Lv Q. Clopidogrel-associated genetic variants on inhibition of platelet activity and clinical outcome for acute coronary syndrome patients. Basic Clin Pharmacol Toxicol. 2019; 124:84-93.
5. Chauhan M, Kumar S, Biswas A. A comprehensive review of the advancement in omic technologies in the field of drug discovery and development. Lett Drug Des Discov. 2024; 21:3319-3331.
6. van den Broek WWA, Ingraham BS, Pereira NL, Lee CR, Cavallari LH, Swen JJ, Angiolillo DJ, Ten Berg JM. Genotype-Guided Antiplatelet Therapy: JACC Review Topic of the Week. J Am Coll Cardiol. 2024; 84:1107-1118.
7. Dong M, Wang L, Zhang Y. Integration of multi-omics approaches in exploring intra-tumoral heterogeneity. Cancer Cell Int. 2025; 25:317.
8. Huang S, Shi D, Dai S, Jiang X, Wang R, Yang M. RNF31 induces paclitaxel resistance by sustaining ALYREF cytoplasmic-nuclear shuttling in human triple-negative breast cancer. Clin Transl Med. 2025; 15:e70203.
9. Ktena I, Willsch D, Böhle M. Challenges and opportunities in multimodal prediction for clinical decision support. Nat Mach Intell. 2024; 6:1425-1435.
10. Zheng L, Zhang S, Li Y, Liu Y, Ge Q, Gu L, Xie Y, Wang X, Ma Y, Liu J, Lu M, Chen Y, Zhu Y, Liu H. MTGNN: a drug-target-disease triplet association prediction model based on multimodal heterogeneous graph neural networks and direction-aware metapaths. J Chem Inf Model. 2025; 65:5921-5933.
11. Liu WZ, Lu PL, Chen YT. MMT-fgMDI: Metapath-driven multimodal transformer framework for predicting fine-grained metabolite-drug interactions. Knowl Based Syst. 2026; 332:114899.
12. Partin A, Brettin TS, Zhu YT, Overbeek J, *et al.* Systematic evaluation and comparison of drug response prediction models: a case study of prediction generalization across cell lines datasets. Cancer Res. 2023; 83(7 Suppl):Abstract 5380.
13. Partin A, Brettin TS, Zhu YT, Shukla M, Xia FF, Yoo HS, Dolezal JM, Kochanny S, Pearson AT, Evrard YA, Doroshow JH, Stevens RL. Drug response prediction in patient-derived xenografts with data augmentation and multimodal deep learning. J Clin Oncol. 2022; 40(16 Suppl):e13572.
14. Ohnuki Y, Akiyama M, Sakakibara Y. Deep learning of multimodal networks with topological regularization for drug repositioning. J Cheminform. 2024; 16:103.
15. Xia XQ, Zhu CY, Zhong F, Liu Y. MDTips: a multimodal-data-based drug-target interaction prediction system fusing knowledge, gene expression profile, and structural data.

Bioinformatics. 2023; 39:btad411.
16. Li Y, Ding JX, Du FQ. Deep learning-based multimodal prediction of NAC response in LARC by integrating MRI and proteomics. Cancer Res Treat. 2025. doi: 10.4143/crt.2025.707
17. Chen G, Sun KX. Leveraging multimodal learning for enhanced drug-target interaction prediction. Front Pharmacol. 2025; 16:1639979.
18. Hua Y, Feng ZH, Song XN, Wu XJ, Kittler J. MMDG-DTI: Drug-target interaction prediction *via* multimodal feature fusion and domain generalization. Pattern Recognit. 2025; 157:110887.
19. Velez-Arce A, Li MM, Gao W, Lin X, Huang K, Fu T. Signals in the cells: multimodal and contextualized machine learning foundations for therapeutics. bioRxiv. 2024:2024.06.12.598655.
20. Chen Y, Zhang Y, Niu M, Wang J, Ren Z, Zou Q. UniSyn: a multi-modal framework with knowledge transfer for anti-cancer drug synergy prediction. Genome Biol. 2026; 27:74.
21. Sederman C, Di Sera T, Qiao Y, *et al*. ScreenDL: a transfer learning framework integrating tumor omics and functional drug screening for personalized clinical drug response prediction. Cancer Res. 2024; 84(6 Suppl):Abstract 907.
22. Zhou Y, Dai Q, Xu Y, Wu S, Cheng M, Zhao B. PharmaFormer predicts clinical drug responses through transfer learning guided by patient derived organoid. NPJ Precis Oncol. 2025; 9:282.
23. Yu J, Shi C, Zhou Y, Liu N, Zong X, Liu Z, Zhang L. DTLCDR: a target-based multimodal fusion deep learning framework for cancer drug response prediction. J Pharm Anal. 2025; 15:101315.
24. Xia XQ, Zhu C, Zhong F, Liu L. TransCDR: a deep learning model for enhancing the generalizability of drug activity prediction through transfer learning and multimodal data fusion. BMC Biol. 2024; 22:227.
25. Rintala TJ, Napolitano F, Fortino V. Multi-task deep latent spaces for cancer survival and drug sensitivity prediction. Bioinformatics. 2024; 40(Suppl 2):ii182-ii189.
26. Xie B, Zhang C, Yan Q. Deep transfer learning in clear cell renal cell carcinoma multi-omics unveils a novel tyrosine kinase inhibitor-tolerant cell subgroup for precision therapy. Res Sq [Preprint]. 2025.
27. Qi G, Wang L, Nan M, Fu Y, Ren Q, Zhang Y, Sun Z, Shi Z, Lu J, Gao J. scXDR: drug response prediction across single-cell datasets *via* heterogeneous network transfer learning. Commun Biol. 2026; 9:139.
28. Zhai J, Liu H. Cross-domain feature disentanglement for interpretable modeling of tumor microenvironment impact on drug response. IEEE J Biomed Health Inform. 2024; 28:4382-4392.
29. Pearl J, Bareinboim E. Generalizing experimental results by leveraging knowledge of mechanisms. Eur J Epidemiol. 2021; 36:149-164.
30. Mera A, Vogt M, Bajorath J. A meta-learning framework to mitigate negative transfer in transfer learning applicable to drug design. Sci Rep. 2025; 15:35236.
31. Chen C, Zhou Y, Li Y. MRDDA: a multi-relational graph neural network for drug-disease association prediction. J Transl Med. 2025; 23:753.
32. Zhang J, Xiong S, Xu Y, Zhang Y. Prediction of cancer drug response based on heterogeneous graph neural networks and multi-omics data. Neural Netw. 2026; 193:108001.
33. Li Y, Shen Y, Cai Y, Zhang Y, Gao J, Huang L. Integrating transcriptomic data with a novel drug efficacy prediction model for TCM active compound discovery. Sci Rep. 2025; 15:7688.
34. Chen M, Zhang M, Yan G, Wang G, Qu C. MRHGNN: enhanced multimodal relational hypergraph neural network for synergistic drug combination forecasting. IEEE Trans Neural Netw Learn Syst. 2025; 36:17086-17098.
35. Yang Y, Li P. GPDRP: a multimodal framework for drug response prediction with graph transformer. BMC Bioinformatics. 2023; 24:484.
36. Chithrananda S, Grand G, Ramsundar B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. In: Proceedings of the Machine Learning for Molecules Workshop, NeurIPS; 2020.
37. Inoue Y, Fu T, Luna A. GraphPINE: Graph Importance Propagation for Interpretable Drug Response Prediction. Bioinformatics. 2026; 42:btad745.
38. Deng L, Yan S, Yang J, Li M, Ding P. GLA-synergy: an interpretable global-local adaptive framework for drug synergy prediction in cancer treatment. J Chem Inf Model. 2025; 65:11385-11399.
39. Zhang H, Sun X, Wang J, Li M, Tang J. A causal inference framework for identifying essential genes to enhance drug synergy prediction. Bioinformatics. 2026; 42:btag010.
40. Tan H, Ji X, Xu CZ, Zhao X, Hou J, Liu M, Ren Y. AGRL-DSE: adaptive graph representation learning on a heterogeneous graph for drug side effect prediction. ACS Omega. 2025; 10:38753-38765.
41. Bian J, Lu H, Wei L, Li Y, Wang G. Relational similarity-based graph contrastive learning for DTI prediction. Brief Bioinform. 2025; 26:bbaf122.
42. Li Y, Hou LX, Yi HC. MOLGAECL: molecular graph contrastive learning *via* graph auto-encoder pretraining and fine-tuning based on drug-drug interaction prediction. J Chem Inf Model. 2025; 65:3104-3116.
43. Chen Y, Huang J, Liu C, Zhang S, Li X, Zhang Z. DualPG-DTA: a large language model-powered graph neural network framework for enhanced drug-target affinity prediction and discovery of novel CDK9 inhibitors exhibiting *in vivo* anti-leukemia activity. Adv Sci (Weinh). 2026; e13099.
44. Liu Z, Lyu K, Li Y, Sun X, Gao X, Yu B. MTEGDRP: interpretable molecular self-attention transformer and equivariant graph neural network based on multi-omics fusion for drug response prediction in cancer cell lines. J Med Chem. 2026; 69:4771-4788.
45. Chen Q, Nie G, Li X, Xu Y, Tan CSH. BioFusionDTI: assimilating graph and sequence modalities for generalizable drug-target interaction prediction. J Chem Inf Model. 2025; 65:13187-13202.
46. Xu W, Liu Y, Li Z. UAMRL: multi-granularity uncertainty-aware multimodal representation learning for drug-target affinity prediction. Bioinformatics. 2025; 41:btaf512.
47. Peng L, Liu X, Chen M, Liao W, Mao J, Zhou L. MGNDTI: A Drug-Target Interaction Prediction Framework Based on Multimodal Representation Learning and the Gating Mechanism. J Chem Inf Model. 2024; 64:6684-6698.
48. Teoh JR, Sarker IH, Sultana N. Advancing healthcare through multimodal data fusion: a comprehensive review of techniques and applications. PeerJ Comput Sci. 2024; 10:e2298.

49. Chu X, Ye Y, Tang S, Han M, Wang G, Lin S. Personalized medication for chronic diseases using multimodal data-driven chain-of-decisions. Adv Sci (Weinh). 2025; 12:e04079.
50. Chou WC, Lin Z. Machine learning and artificial intelligence in physiologically based pharmacokinetic modeling. Toxicol Sci. 2023; 191:1-14.
51. Tanihata S, Iwata H. Next-Generation Artificial Intelligence for ADME Prediction in Drug Discovery: From Small Molecules to Biologics. Yonago Acta Med. 2026; 69:1-13.
52. Lu J, Deng K, Zhang X, Liu G, Guan Y. Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. iScience. 2021; 24:102804.
53. Giacometti T, Rocchi E, Cojutti PG, Magnani F, Remondini D, Pea F, Castellani G. Leveraging Neural ODEs for Population Pharmacokinetics of Dalbavancin in Sparse Clinical Data. Entropy (Basel). 2025; 27:602.
54. Lu J, Bender B, Jin JY, Guan Y. Deep learning prediction of patient response time course from early data *via* neural-pharmacokinetic/pharmacodynamic modelling. Nat Mach Intell. 2021; 3:696-704.
55. Bräm DS, Nahum U, Schropp J, Pfister M, Koch G. Low-dimensional neural ODEs and their application in pharmacokinetics. J Pharmacokinet Pharmacodyn. 2024; 51:123-140.
56. Fonseca LL, Laubenbacher RC, Böttcher L. Learning dynamical systems with biochemically informed neural ordinary differential equations. bioRxiv. 2026.
57. Ahmadi N, Cao Q, Humphrey JD, Karniadakis GE. Physics-Informed Machine Learning in Biomedical Science and Engineering. Annu Rev Biomed Eng. 2026; 28:309-336.
58. Lee K, Cho D, Jang J, Choi K, Jeong HO, Seo J, Jeong WK, Lee S. RAMP: response-aware multi-task learning with contrastive regularization for cancer drug response prediction. Brief Bioinform. 2023; 24:bbac504.
59. Li YY, Chen L, Pu C, Zang CD, Yan YC, Chen YC, Zhang YM, Liu HC. Co-model for chemical toxicity prediction based on multi-task deep learning. Mol Inform. 2023; 42:e2200257.
60. Sharma B, Chenthamarakshan V, Dhurandhar A, Pereira S, Hendler JA, Dordick JS, Das P. Accurate clinical toxicity prediction using multi-task deep neural nets and contrastive molecular explanations. Sci Rep. 2023; 13:4908.
61. Gu Y, Yu Z, Wang Y, Chen L, Lou C, Yang C, Li W, Liu G, Tang Y. admetSAR3.0: a comprehensive platform for exploration, prediction and optimization of chemical ADMET properties. Nucleic Acids Res. 2024; 52:W432-W438.
62. Cheng Y, Hu H, Dong X, Hao X, Li Y. Exploring Transformer Model in Longitudinal Pharmacokinetic/Pharmacodynamic Analyses and Comparing with Alternative Natural Language Processing Models. J Pharm Sci. 2024; 113:1368-1375.
63. Li X, Zhan M, Fang J, Liu G, Tang Y, Li W. MMPK: A Multimodal Deep Learning Framework to Predict Human Oral Pharmacokinetic Parameters. J Med Chem. 2025; 68:16678-16690.
64. Cui Y, Ji X, Guo W, Chen S, Shen T, Chen L, Ke G, Jin C, Gao Z, Sun W. Toward Generalizable Data-Driven Pharmacokinetics with Interpretable Neural ODEs. J Chem Inf Model. 2026; 66:2640-2650.
65. Chen Y, Zhang Y, Niu M, Wang J, Ren Z, Zou Q, Song J, Luo X. Hierarchical multi-omics data integration and modeling predict cell-specific chemical proteomics and drug responses. Cell Rep Methods. 2023; 3:100452.
66. Chen Y, Zhang Y, Niu M, Wang J, Ren Z, Zou Q, Song J, Luo X. UniSyn: a multi-modal framework with knowledge transfer for anti-cancer drug synergy prediction. Genome Biol. 2026; 27:74.
67. CA H, AS J. PerturbSynX: Deep learning framework for predicting drug combination synergy scores using drug induced gene perturbation data. Comput Biol Chem. 2026; 120:108736.
68. Ota R, Yamashita F. Application of machine learning techniques to the analysis and prediction of drug pharmacokinetics. J Control Release. 2022; 352:961-969.
69. Peng W, Xu X, Lin J, Chen G, Dai W, Fu X, Li L, Liu L, Yu N. Predicting Anti-Cancer Drug Response Based on Hypergraph Representation Learning. IEEE Trans Comput Biol Bioinform. 2025; 22:2430-2441.
70. Ma H, Wang Z, Jia Y, Li Z, Zhao Y, Yu C. Prediction of Cancer Drug Response Based on Hypergraph Convolutional Network and Contrastive Learning. IEEE Trans Comput Biol Bioinform. 2026; 23:1463-1474.
71. Chen Y, Zhang L. Hi-GeoMVP: a hierarchical geometry-enhanced deep learning model for drug response prediction. Bioinformatics. 2024; 40:btae204.
72. Bräm DS, Steiert B, Pfister M, Steffens B, Koch G. Low-dimensional neural ordinary differential equations accounting for inter-individual variability implemented in Monolix and NONMEM. CPT Pharmacometrics Syst Pharmacol. 2025; 14:5-16.
73. Bräm DS, Steiert B, Steffens B, Pfister M, Koch G. Automated Pharmacometric Model Development by Leveraging Low-Dimensional Neural ODEs and LASSO Regression. CPT Pharmacometrics Syst Pharmacol. 2026; 15:e70285.
74. Gui CP, Chen YH, Zhao HW, *et al*. Multimodal recurrence scoring system for prediction of clear cell renal cell carcinoma outcome: a discovery and validation study. Lancet Digit Health. 2023; 5:e515-e524.
75. Zheng QY, Mei HN, Weng XD, Yang R, Jiao PP, Ni XM, Yang XX, Wu JJ, Fan J, Yuan JP, Liu XH, Chen ZY. Artificial intelligence-based multimodal prediction for nuclear grading status and prognosis of clear cell renal cell carcinoma: a multicenter cohort study. Int J Surg. 2025; 111:3722-3730.
76. Arslan S, Schmidt J, Bass C, Mehrotra D, Geraldes A, Singhal S, Hense J, Li X, Raharja-Liu P, Maiques O, Kather JN, Pandya P. A systematic pan-cancer study on deep learning-based prediction of multi-omic biomarkers from routine pathology images. Commun Med (Lond). 2024; 4:48.
77. Partin A, Brettin T, Zhu Y, Dolezal JM, Kochanny S, Pearson AT, Shukla M, Evrard YA, Doroshow JH, Stevens RL. Data augmentation and multimodal learning for predicting drug response in patient-derived xenografts from gene expressions and histology images. Front Med (Lausanne). 2023; 10:1058919.
78. Saeed D, Xing H, Feng L. A scalable multimodal graph neural network for drug combination response prediction. Mol Divers. 2026; 30:1695-1708.
79. Luo T, Zhang Z, Chen XG, Li Z. Multi-modal contrastive drug synergy prediction model guided by single modality. J Cheminform. 2025; 17:146.
80. Li ML, Cao C, Zou Q, Wei LY, Wang YS. MHAFR-

DDI: a multimodal hierarchical attention fusion and relation-aware architecture for drug-drug interaction event prediction. Brief Bioinform. 2026; 27:bbag077.
81. Zhang L, Kang X, Yang X, Wang L, Yang G, Chu J. MOSAIC: a multi-granularity cross-modal framework for predicting synergistic drug combinations in personalized healthcare. IEEE J Biomed Health Inform. 2025. doi: 10.1109/JBHI.2025.3605313.
82. Feng YH, Feng ZL, Yan XY, Zhang SW, Shi JY. Few-shot drug synergy prediction *via* rapid cross-tier adaptation meta-optimization. Brief Bioinform. 2025; 26:bbaf683.
83. Zeng P, Zhang B, Liu A, Meng Y, Tang X, Yang J, Xu J. Drug repositioning based on tripartite cross-network embedding and graph convolutional network. Expert Syst Appl. 2024; 252:124152.
84. Gunther K, Liu D, Stannard G, Holmes M, Loo C, Guo B, Bowden N, Abbott J, Ford CE. An integrated multimodal approach to drug repurposing in endometriosis, using ROR1 as a target. Front Pharmacol. 2025; 16:1716062.
85. Qiu Y, Cheng F. Artificial intelligence for drug discovery and development in Alzheimer's disease. Curr Opin Struct Biol. 2024; 85:102776.
86. Xu J, Mao C, Hou Y, *et al.* Interpretable deep learning translation of GWAS and multi-omics findings to identify pathobiology and drug repurposing in Alzheimer's disease. Cell Rep. 2022; 41:111717.
87. Mijit A, Zhang YZ, Wu Y, Li M, Wang Z, Cheng LF. Dual inhibition of EGR1/STAT3 transcriptional hubs suppresses macrophage-driven liver fibrosis: a multi-omics-guided drug repurposing strategy. Biochem Pharmacol. 2025; 240:117120.
88. Hwang W, Han N. Identification of potential pan-coronavirus therapies using a computational drug repurposing platform. Methods. 2022; 203:214-225.
89. Xie W, Yu J, Huang L, For LS, Zheng Z, Chen X, Wang Y, Liu Z, Peng C, Wong KC. DeepSeq2Drug: a multimodal deep learning framework with transfer learning for antiviral drug repurposing. Comput Biol Med. 2024; 175:108487.
90. Pun FW, Leung GHD, Leung HW, Liu BHM, Long X, Ozerov IV, Wang J, Ren F, Aliper A, Izumchenko E, Moskalev A, de Magalhães JP, Zhavoronkov A. Hallmarks of aging-based dual-purpose disease and age-associated targets predicted using PandaOmics AI-powered discovery engine. Aging (Albany NY). 2022; 14:2475-2506.
91. Wu Y, Huang J, Guo J, Lian W, Suo M. Drug repurposing in traditional Chinese medicine: from empirical wisdom to modern therapeutic strategies. Front Pharmacol. 2025; 16:1631727.
92. Sahu M, Gupta R, Ambasta RK, Kumar P. Artificial intelligence and machine learning in precision medicine: a paradigm shift in big data analysis. Prog Mol Biol Transl Sci. 2022; 190:57-100.
93. Singh Y, Andersen JB, Hathaway QA, Vera-Garcia DV, Keishing V, Venkatesh SK, Salehi S, Povero D, Wallace MB, Gores GJ, Wei Y, Horvat N, Erickson BJ, Quaia E. Leveraging multimodal foundation models in biliary tract cancer research. Tomography. 2025; 11:96.
94. Ren Y, Pieper AA, Cheng F. Utilization of precision medicine digital twins for drug discovery in Alzheimer's disease. Neurotherapeutics. 2025; 22:e00553.
95. Gonzalez G, Lin X, Herath I, Veselkov K, Bronstein M, Zitnik M. Combinatorial prediction of therapeutic perturbations using causally inspired neural networks. Nat Biomed Eng. 2026; 10:1008-1025.
96. Yang H, Tong X. MDEbert-DTA: drug-target affinity prediction using multimodal contrastive learning and dynamic graph aggregation. Appl Intell. 2026; 56:191.
97. Cong Y, Endo T. Multi-omics and artificial intelligence-guided drug repositioning: prospects, challenges, and lessons learned from COVID-19. OMICS. 2022; 26:361-371.
98. Fu C, Chen Q. The future of pharmaceuticals: artificial intelligence in drug discovery and development. J Pharm Anal. 2025; 15:101248.
99. Li Z, Tu X, Chen Y, Lin W. HetDDI: a pre-trained heterogeneous graph neural network model for drug-drug interaction prediction. Brief Bioinform. 2023; 24:bbad385.
100. Sun C, Man XF, Li N, Liu ZH, Chen Z, Yu B. FG-PromptNet: cross-modal prompt learning for explainable molecular property prediction. Comput Biol Chem. 2026; 124:109096.

Received June 15, 2026; Revised August 9, 2026; Accepted August 15, 2026.

**Address correspondence to:*
Qichao Luo, School of Pharmacy, Anhui Medical University, Hefei 230032, Anhui, China.
E-mail: luoqichao@ahmu.edu.cn

Released online in J-STAGE as advance publication August 23, 2026.