Multi-Modal Deep Learning for Accurate Drug-Target Interaction Prediction in Precision Medicine

[1]Siyuan Zhang , [2]Maya V. Sharma , [3]Wei Liu
School of Computer Science, Ludong University, Yantai, Shandong, China

ABSTRACT

Precision medicine requires accurate prediction of drug–target interactions (DTIs) to enable individualized therapeutic selection and minimize adverse effects. This review synthesizes multi-modal deep learning frameworks for DTI prediction, positioning them as exemplar intelligent engineering systems (IES) empowered by artificial intelligence. Grounded in cyber-physical systems (CPS) theory and information theory, these frameworks integrate heterogeneous data modalities—molecular graphs, protein sequences, gene expression profiles, and clinical records—through graph neural networks, transformers, and cross-modal fusion architectures. The analysis traces four AI-driven transformation pathways: from population-level standardized screening to personalized, flexible prediction models; from reactive post-hoc validation to proactive interaction forecasting; from manual expert curation to autonomous intelligent inference; and from isolated computational tools to integrated cyber-physical ecosystems linking research, clinical decision support, and patient monitoring. Concrete examples include multi-modal transformer networks processing SMILES strings and amino acid sequences for binding affinity regression, and graph-based models fused with electronic health record data for adverse-event prediction. While these approaches substantially improve predictive accuracy and generalization to unseen compounds, critical limitations persist regarding data heterogeneity, model interpretability, and distribution shift across patient populations. Future prospects center on physics-informed multi-modal architectures, federated learning for privacy-preserving integration, and digital-twin-enabled virtual screening. The work provides a rigorous template for advancing trustworthy AI in precision medicine within broader IES paradigms.

Keywords: Drug-Target Interaction Prediction; Multi-Modal Deep Learning; Precision Medicine; Cyber-Physical Systems; Graph Neural Networks; Transformer Fusion; Personalized Therapeutics; Intelligent Engineering Systems

1. INTRODUCTION

The promise of precision medicine hinges on the ability to predict how individual patients will respond to specific drugs, a challenge centrally addressed by accurate drug–target interaction (DTI) modeling. Traditional experimental screening is time-consuming, costly, and limited in throughput, while single-modality computational methods frequently fail to capture the complex, multi-scale biology governing binding affinity and downstream phenotypic response [1]. Multi-modal deep learning has emerged as a powerful paradigm that jointly processes heterogeneous data sources—chemical structures represented as graphs or SMILES strings, protein sequences or 3D structures, transcriptomic profiles, and clinical covariates—to learn richer representations and achieve superior predictive performance on both known and novel drug–target pairs.

These frameworks operate as intelligent engineering systems in which biological “plants” (molecular interactions and patient physiology) are coupled with computational perception and decision layers through continuous data flows. Grounded in systems engineering for holistic pipeline design, cyber-physical systems theory for closed-loop feedback between prediction and clinical outcome monitoring, and information theory for quantifying uncertainty reduction across modalities, multi-modal DTI models exemplify the AI-driven transformation of biomedical engineering systems. This review examines the theoretical foundations, transformation pathways, concrete applications, and forward-looking challenges of such frameworks, maintaining a critical balance between demonstrated gains in accuracy and generalization and persistent risks of overfitting, data bias, and limited interpretability. Concrete real-world illustrations are drawn from validated architectures trained on benchmark datasets such as Davis, KIBA, and BindingDB, as well as early clinical decision-support deployments.

2. The Theoretical Foundation of Artificial Intelligence and Intelligent Engineering Systems

2.1 The Connotation and Characteristics of Intelligent Engineering Systems

Intelligent engineering systems are engineered artifacts that exhibit adaptive, goal-directed behavior through the tight integration of physical or biological processes, computational intelligence, and bidirectional data flows. Core characteristics include autonomy in decision-making under uncertainty, adaptability via continuous learning, resilience through self-monitoring and drift detection, and interconnectivity enabling system-of-systems coordination [2]. These attributes map directly onto cyber-physical architectures in which physical or biological entities are augmented by cyber layers that close real-

time feedback loops.

In the domain of precision medicine, a multi-modal DTI prediction system constitutes a mobile, data-driven IES. Molecular and clinical data streams are fused by graph neural networks and transformer encoders to produce binding-affinity or interaction-probability outputs; these outputs inform therapeutic selection and are subsequently validated against real-world patient outcomes, thereby closing the CPS loop. The system perceives heterogeneous biological signals, reasons about interaction likelihood and potential adverse effects, and supports autonomous or semi-autonomous clinical recommendations. Reviews of multi-modal representation learning for DTI underscore the evolution from unimodal sequence or graph models toward architectures that explicitly model cross-modal dependencies, yielding measurable improvements in generalization to unseen chemical entities and protein targets [1,3].

2.2 The Theoretical Logic of Artificial Intelligence Empowering Engineering Systems

The theoretical logic rests on three interlocking foundations. Systems engineering provides the integrative methodology for requirements decomposition, modality alignment, and end-to-end verification of the perception–fusion–prediction pipeline. CPS theory supplies the architectural pattern of networked computation tightly coupled with biological dynamics, enabling continuous model updating from clinical feedback. Information theory furnishes the quantitative basis: DTI prediction corresponds to reduction of uncertainty (entropy) regarding binding events across heterogeneous data modalities; optimal fusion architectures maximize mutual information between concatenated or cross-attended representations and true interaction labels [3].

AI empowers the engineering system by learning hierarchical, non-linear mappings that traditional molecular docking or single-modality machine learning cannot capture. Graph neural networks encode local chemical environments and topological features of drug molecules, transformers capture long-range dependencies in protein sequences, and cross-modal attention or late-fusion layers integrate these representations for joint prediction. An original insight is that multi-modal DTI frameworks perform progressive “information-to-decision” compression across biological scales: raw multi-omics and chemical data are abstracted into compact latent spaces that preserve interaction-relevant signals while discarding modality-specific noise, thereby enabling both high predictive accuracy and computational tractability for clinical deployment.

3. The Transformation Path of Intelligent Engineering Systems Mode Driven by Artificial Intelligence

3.1 Transition from Standardized Production to Personalized and Flexible Manufacturing

In the context of drug development, standardized “production” historically referred to population-level high-throughput screening and one-size-fits-all efficacy assumptions. Multi-modal deep learning enables a transition to personalized, flexible “manufacturing” of patient-specific DTI predictions. Graph neural networks process drug molecular graphs while transformer encoders handle protein sequences or expression profiles; cross-modal fusion layers then generate individualized binding-affinity estimates conditioned on patient genomic or transcriptomic context [1,4]. Concrete examples include architectures that incorporate gene-expression signatures from the LINCS L1000 project alongside chemical structures to predict drug response in specific cancer subtypes. This personalization improves therapeutic matching and reduces off-target effects, yet introduces risks of overfitting to limited patient-stratified datasets and requires rigorous external validation across diverse ancestries and disease contexts.

3.2 Transition from Reactive Response to Proactive Prediction and Optimization

Traditional DTI workflows remain largely reactive: interactions are characterized experimentally after initial screening or observed clinically after market approval. Multi-modal frameworks shift toward proactive prediction and optimization. Multi-task or multi-modal models jointly predict binding affinity, selectivity, and potential adverse-event profiles by fusing chemical, proteomic, and clinical modalities [3,5]. Reinforcement learning can further optimize multi-drug regimens or dosing schedules by treating predicted interaction landscapes as reward signals. Real-world illustrations include hybrid models that forecast both on-target efficacy and off-target toxicity for kinase inhibitors across patient-derived xenograft models, enabling earlier de-risking of development candidates. The proactive stance accelerates discovery timelines and improves safety margins, although prediction errors on novel scaffolds or underrepresented patient populations can propagate costly downstream failures.

3.3 Transition from Manual Operation to Autonomous Intelligent Control

Expert curation of literature, manual feature engineering, and rule-based filtering previously dominated DTI workflows. Multi-modal deep learning enables a transition to autonomous intelligent control through end-to-end or hierarchical pipelines. Graph neural networks and transformers operating on edge or cloud platforms generate real-time interaction predictions that feed directly into clinical decision-support systems or automated virtual screening campaigns [4,6]. Edge deployment on hospital servers or secure cloud instances supports low-latency inference for precision oncology boards, while federated learning

variants preserve patient privacy. Demonstrated systems achieve competitive or superior performance to expert-curated baselines on benchmark DTI datasets while scaling to millions of compound–target pairs. Autonomy reduces human bottleneck and bias, yet demands robust uncertainty quantification and human-in-the-loop oversight for high-stakes therapeutic decisions.

3.4 Transition from Isolated Systems to Integrated Cyber-Physical Ecosystems

Isolated computational models limit DTI prediction to static datasets and preclude continuous learning from real-world outcomes. Integration into cyber-physical ecosystems connects predictive engines to electronic health records, pharmacovigilance databases, wearable sensor streams, and digital patient twins [5,7]. Shared representations enable population-level model updating while individual predictions remain personalized. Digital twins of patients—virtual replicas updated with longitudinal multi-omics and clinical data—support in silico simulation of drug combinations and response trajectories before physical administration. This ecosystem perspective transforms DTI prediction from a standalone bioinformatics task into a resilient node within learning healthcare systems, while elevating data privacy, interoperability standards, and model governance to first-order design requirements.

4. Applications of Artificial Intelligence in Intelligent Engineering Systems

4.1 Application in Intelligent Fault Diagnosis and Predictive Maintenance

Data-driven fault diagnosis and predictive maintenance enhance IES resilience in clinical AI pipelines. Concept drift detectors and anomaly detection models monitor shifts in input distributions (new drug scaffolds, evolving patient demographics) or degradation in predictive calibration on DTI tasks [5,8]. Digital twins of the prediction model itself track performance metrics across deployment sites and trigger retraining or recalibration when drift exceeds thresholds. Concrete examples include monitoring frameworks that detect performance decay on underrepresented ethnic groups or novel therapeutic classes and automatically flag the need for targeted data augmentation or modality re-weighting. This approach maintains long-term reliability of DTI predictions in dynamic clinical environments, although it requires high-quality labeled outcome data and careful management of false-positive drift alerts that could disrupt care workflows.

4.2 Application in Intelligent Process Optimization and Quality Control

Treating the multi-modal DTI pipeline as the optimizable “process” yields substantial gains in predictive quality. Multi-objective optimization and neural architecture search tune fusion strategies, modality weights, and loss functions to balance accuracy, calibration, and computational cost across heterogeneous datasets [3,6]. Quality control is enforced through rigorous benchmarking on held-out chemical and biological spaces, temporal validation splits that mimic real-world deployment, and prospective clinical studies. Computer-vision or graph-based modules provide real-time “in-line inspection” of molecular and proteomic features, with continuous learning from logged prediction–outcome pairs. These techniques have produced documented improvements in both binding-affinity regression metrics (e.g., concordance index) and downstream clinical utility indicators, although certifying generalization across the full space of possible drug–target pairs remains an open challenge.

4.3 Application in Intelligent Human-Machine Collaboration and Operational Efficiency Enhancement

Full autonomy in therapeutic decision-making is neither ethically desirable nor currently achievable. Intelligent human–machine collaboration maximizes efficiency while preserving clinician accountability. Multi-modal DTI systems provide decision support through ranked interaction predictions, confidence intervals, and mechanistic explanations derived from attention weights or integrated gradients [4,7]. Digital-twin dashboards enable oncologists or pharmacologists to simulate “what-if” scenarios for alternative compounds or combinations before final selection. This collaborative model accelerates precision oncology workflows and reduces cognitive load, allowing specialists to focus on complex cases and patient communication. Interface design must actively mitigate over-reliance on model outputs and clearly communicate uncertainty and limitations to preserve appropriate human oversight.

5. Challenges and Prospects of Artificial Intelligence Empowering Intelligent Engineering Systems

5.1 Current Main Challenges

Heterogeneous data quality and missing modalities remain pervasive: many patient records lack complete genomic or proteomic profiling, and chemical libraries are biased toward historical drug space [1,5]. Distribution shift across institutions, ancestries, and evolving disease landscapes degrades model performance in prospective use. Interpretability of cross-modal fusion decisions is limited, complicating regulatory acceptance and clinician trust. Privacy constraints and data governance requirements hinder large-scale multi-institutional training, while adversarial or spurious correlations in multi-omics data can produce over-optimistic in silico performance that fails to translate clinically. Finally, rigorous prospective validation and health-economic evaluation of multi-modal DTI systems lag behind technical development, slowing responsible translation.

5.2 Prospects for Future Development Trends

Promising trajectories address these limitations through physics-informed and biologically grounded multi-modal architectures that embed structural or mechanistic priors into fusion layers, improving both accuracy and interpretability [6,8]. Federated and split learning paradigms enable privacy-preserving collaborative training across hospitals without centralizing sensitive data. Digital-twin ecosystems will mature from descriptive patient replicas to prescriptive platforms that simulate individualized DTI landscapes and optimize combination therapies in silico. Advances in explainable AI, uncertainty quantification, and causal representation learning will support regulatory-grade evidence generation. Over the medium term, these developments are expected to embed multi-modal DTI prediction as a routine, trustworthy component of learning healthcare systems, accelerating the realization of truly personalized therapeutics while furnishing transferable principles for IES development in other data-rich biomedical domains.

6. Conclusion

This review has articulated multi-modal deep learning frameworks for drug–target interaction prediction as a paradigmatic manifestation of AI-empowered intelligent engineering systems in precision medicine. By grounding the analysis in systems engineering, cyber-physical systems theory, and information theory, the framework illuminates four interlocking transformation pathways that shift drug discovery and therapeutic decision-making from reactive, population-level, and isolated modes toward proactive, personalized, and ecosystem-integrated intelligence. Applications in model monitoring, process optimization, and human–AI collaboration demonstrate concrete benefits in predictive accuracy, clinical efficiency, and resilience, supported by benchmark and early translational evidence [1–8]. Persistent challenges in data heterogeneity, generalization, interpretability, and prospective validation nevertheless require sustained interdisciplinary effort. Future progress hinges on biologically informed architectures, privacy-preserving collaboration, and rigorous assurance methodologies capable of delivering certifiable, equitable, and clinically actionable DTI predictions. Successful realization will materially advance precision medicine while offering generalizable insights for the design of trustworthy intelligent engineering systems across healthcare and beyond.

REFERENCES

[1] Zhou, D., et al. (2021). MultiDTI: drug–target interaction prediction based on multi-modal representation learning to bridge the gap between new chemical entities and known ligands. Bioinformatics, 37(23), 4485–4492. https://doi.org/10.1093/bioinformatics/btab448

[2] Vagale, A., et al. (2021). Path planning and collision avoidance for autonomous surface vehicles I: a review. Journal of Marine Science and Technology, 26, 1292–1306. https://doi.org/10.1007/s00773-020-00787-6

[3] Sarhadi, P., Naeem, W., & Athanasopoulos, N. (2022). A survey of recent machine learning solutions for ship collision avoidance and mission planning. IFAC-PapersOnLine, 55(31), 257–268. https://doi.org/10.1016/j.ifacol.2022.10.286

[4] Ma, Y., et al. (2020). Collision-avoidance under COLREGS for unmanned surface vehicles via deep reinforcement learning. Maritime Policy & Management, 47(5), 665–686. https://doi.org/10.1080/03088839.2020.1756494

[5] Xing, B., et al. (2023). A review of path planning for unmanned surface vehicles. Journal of Marine Science and Engineering, 11(8), 1556. https://doi.org/10.3390/jmse11081556

[6] Wang, W., et al. (2022). A COLREGs-compliant collision avoidance decision approach based on deep reinforcement learning. Journal of Marine Science and Engineering, 10(7), 944. https://doi.org/10.3390/jmse10070944

[7] Fan, Y., et al. (2022). A novel reinforcement learning collision avoidance algorithm for USV. Journal of Marine Science and Engineering, 10(3), 346. https://doi.org/10.3390/jmse10030346

[8] Burmeister, H. C., et al. (2021). Autonomous collision avoidance at sea: A survey. Journal of Marine Science and Engineering, 9(9), 1027. https://doi.org/10.3390/jmse9091027