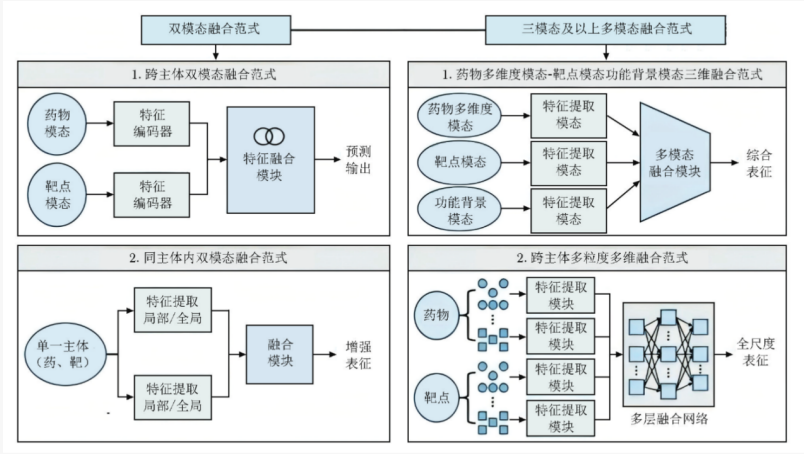
AI maps a clearer path to drug-target discovery

GA, UNITED STATES, August 20, 2026 /EINPresswire.com/ -- Artificial intelligence is reshaping one of drug discovery's most consequential questions: which compounds interact with which biological targets, and how strongly. A new review maps the fast-moving landscape of artificial intelligence (AI)-driven drug-target interaction (DTI) prediction, showing how the field has progressed from hand-crafted molecular features to deep learning, graph-based models, pretraining and multimodal fusion. The authors organize the review around data foundations, model architectures, benchmark resources and evaluation practices that determine whether predictions are trustworthy. They also argue that future progress will depend less on simply building larger models and more on improving data quality, cross-dataset robustness, mechanistic interpretability and links to experimental testing.



Multimodal-based methods for DTI prediction

Drug-target interaction prediction can help researchers explain mechanisms of action, identify candidate targets, reposition existing medicines and prioritize compounds before costly laboratory work. Yet the field still faces uneven structural coverage, scarce high-quality negative examples, inconsistent activity labels and strong bias toward well-studied drugs and targets. Models that perform well on familiar benchmark datasets may lose accuracy when asked to evaluate a new compound, a new protein family or data collected under different experimental conditions. Multimodal systems can add chemical, structural, functional and network context, but they also introduce alignment problems, higher computational demands and fresh sources of noise. Because of these challenges, deeper investigation is needed into standardized dataset construction, data quality control, realistic evaluation and experimentally testable prediction frameworks.

Published online on July 20, 2026, in the Medical Journal of Peking Union Medical College Hospital, the review was prepared by researchers from the School of Artificial Intelligence at Beijing University of Posts and Telecommunications and the Shandong Computer Science

Center, also known as the National Supercomputing Center in Jinan. The team examines the data foundations, benchmark resources, modeling strategies and translational barriers shaping artificial intelligence (AI)-driven drug-target interaction (DTI) prediction, while outlining a path from algorithm-centered comparisons toward predictive frameworks that support real-world drug discovery workflows, pharmacological research and preclinical validation.

The review begins with three core data types: drug representations, target representations and drug-target association labels. Drugs may be encoded through the simplified molecular input line entry system (SMILES), molecular fingerprints or molecular graphs, while targets can be described through amino-acid sequences, three-dimensional structures and functional annotations. The authors then trace three stages of technical development. Traditional machine learning approaches include similarity-based methods, matrix factorization, network-based models, engineered-feature models, and hybrid models that combine multiple strategies ; deep learning automatically extracts local sequence patterns, molecular topology, long-range dependencies, semantic representations learned through large-scale pretraining, and enhanced signals from external knowledge or higher-order structural priors; and multimodal models integrate multi-source information, including drug, protein, disease, side-effect, perturbation phenotype, and knowledge-network data. Graph neural networks (GNNs), attention mechanisms, Transformer architectures and pretrained representations have expanded what models can capture, especially when molecular structure and biological context must be considered together. The review also compares benchmark resources and shows that each serves a different task, from binary interaction classification to quantitative affinity prediction and structure-based virtual screening. However, performance cannot be compared fairly without consistent datasets, split protocols and metrics. Random splits may exaggerate success by placing highly similar examples in training and test sets. More realistic cross-distribution, target-family and cold-start evaluations are needed to reveal whether a model can generalize beyond familiar chemical and biological space.

The authors said the next advance should not be measured only by a higher score on a benchmark. They said useful DTI systems must show where their evidence comes from, remain reliable when data distributions change and generate hypotheses that researchers can test. In their view, predictions become more valuable when they narrow candidate lists, suggest plausible binding mechanisms and guide molecular docking, structural modeling and laboratory experiments. They also emphasized that adding more data types is not automatically beneficial: each modality must be relevant, well aligned and strong enough to improve biological reasoning rather than merely increase model complexity.

The framework could support faster candidate screening, drug repurposing, lead optimization and potential risk assessment, but the review treats AI as a decision aid rather than a replacement for experiments. A credible workflow would connect target biology, computational ranking, structural modeling and staged validation through binding assays, cellular studies and animal research. The authors also foresee specialized biomedical foundation models and intelligent agents that could integrate literature, databases, candidate generation and validation

design within one traceable workflow. For pharmaceutical research, the central implication is practical: AI-driven DTI prediction will have its greatest impact when it shortens the path from large search spaces to a smaller set of biologically justified, experimentally verifiable options.

References

DOI
10.12290/xhyxzz.2026-0399

Original Source URL
https://xhyxzz.pumch.cn/article/doi/10.12290/xhyxzz.2026-0399

Funding information
National Key Research and Development Program of China, grant 2025YFE0216800.

Lucy Wang
BioDesign Research
email us here