Prediction of Drug Target Interactions Based on Artificial Intelligence

Qizhong Yang*

College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

Abstract. Drug-target interaction (DTI) prediction is an efficient pre-screening method that uses algorithmic models to assess the binding potential of drug molecules to protein targets. Current research is accelerating towards the integration of heterogeneous graph neural networks, protein language models, and generative artificial intelligence. This review systematically summarizes the latest developments in these technologies, pointing out the problems currently being addressed in research such as data sparsity and cold start, as well as the manifestations of general machine learning challenges such as recommendation systems and noise learning in the biomedical field; Interpret representative models such as Dual Heterogeneous Graph Transformer for Drug-Target Interaction (DHGT-DTI) and Graph Positional encoding and Sequence features for Drug-Target Interaction (GPS-DTI). The combination of dual perspective learning, equivariant graph convolution, and attention mechanism enhances the understanding and reasoning ability of these models in complex biological networks. The generative artificial intelligence diffusion model has opened up a path for developing new drugs through structured and data enhanced approaches. Research has shown that important issues related to computational performance, interpretability, and data compatibility still need to be addressed in existing models. Building a high-performance, multifunctional pre trained model for large-scale biomolecules should be a key direction for development.

1 Introduction

Although new drug development has long been plagued by high prices, long cycles, and low success rates, it also limits the effectiveness of traditional laboratory screening methods in exploring the vast chemical space. However, virtual screening techniques based on molecular docking and quantitative structure-activity relationships suffer from limited generalization ability and inability to explore the complex dynamic and nonlinear relationships in Drug-target interaction (DTI) due to the use of artificially constructed molecular descriptors, preset physical force constants, and assumptions about the three-dimensional structure of target proteins.

The development of new drugs using computer-aided design, especially the prediction of DTI, can provide necessary information for drug discovery, reduce the time required for experimental validation, and improve success rates, playing an increasingly important role in drug development; Before conducting experiments, it is possible to directly determine whether small molecule drugs have the ability and affinity to bind to target proteins and other macromolecules. In this way, effective preliminary screening can be conducted for the subsequent experiments, narrowing down the research scope and saving trial and error costs.

The rise of deep learning, especially graph neural networks and transformer architectures, excels at handling non-Euclidean data and can understand sequential context. This provides us with the possibility to directly extract complex features from raw data such as molecular graphs and protein sequences [1]. The essence of drug target binding is a physical and chemical process in three-dimensional space. Geometric deep learning provides a mathematical basis for accurately processing the three-dimensional structural information of molecules and proteins by introducing equivariant neural networks that strictly follow the symmetry of three-dimensional space. The booming development of generative artificial intelligence, especially diffusion models, can not only efficiently generate reasonable three-dimensional conformational expansion data of molecules, but also design new molecules with specific properties from scratch, thus promoting the new stage of DTI research from passive prediction to active design.

The current research in this field has evolved from the application of a single technology to the deep integration and collaborative innovation of four major technological directions: heterogeneous graph neural networks, protein language models, geometric deep learning, and generative artificial intelligence. Graph neural networks and Transformers are fused together to form a hybrid model with stronger expressive power. Multi-modal information from sequences, 2D graphs, 3D coordinates to knowledge graphs is used for collaborative modelling. Discriminant prediction models and generative design models are gradually forming a closed loop from generation to evaluation. However, this technological integration also brings new

* Corresponding author: yangqizhong@cau.edu.cn

challenges including model complexity and computational efficiency, sparsity of biomedical data, noise, and distribution bias.

This review aims to systematically sort out and summarize the current research on DTI prediction based on artificial intelligence. The core goal is to clearly explain the basic principles of the four key technologies mentioned above and their value in DTI prediction, deeply analyze how these technologies can address specific challenges through deep integration, and analyze the core bottlenecks of current advanced models.

2 Drug target interaction prediction based on heterogeneous graph neural network

Heterogeneous graphs can naturally characterize multiple entities such as drugs, targets, diseases, and their complex relationships, making them a powerful framework for constructing DTI prediction models. Recent research is no longer satisfied with the simple application of pre-defined meta paths, but instead pursues higher-order reasoning that is more expressive and adaptive.

Taking the Dual Heterogeneous Graph Transformer for Drug-Target Interaction (DHGT-DTI) model as an example, it innovatively adopts the design of dual view fusion of neighborhood and meta path [2]. From a neighborhood perspective, the model aggregates the features of directly connected nodes through Graph Sample and Aggregate (GraphSAGE) of heterogeneous graph versions to capture local network topology; From the perspective of meta paths, the graph Transformer is used to aggregate the attention of semantic neighbors defined by meta paths, thus modeling high-order functional associations that span multiple steps. The representations of the two perspectives are adaptively combined through a gated attention mechanism. This method clearly distinguishes and combines local connectivity and global semantic relationships in the network, and has better generalization ability when facing new biological pathways or complex disease phenotype prediction tasks.

In order to overcome the problem of imbalanced positive and negative samples in DTI data, a Graph wavelet transform and Hierarchical Contrastive learning for Drug-Target Interaction (GHCDTI) prediction heterogeneous network drug target interaction model was designed based on graph wavelet transform and multi-level contrastive learning. Comparative learning was performed on drug target expression to ensure that the model could distinguish positive and negative samples. Multiple levels of self-supervised contrastive learning tasks were constructed for subgraphs and the entire image to assist in training [3]. The self-supervised signals at different levels can guide the model to learn more robust and effective molecular target description information, reducing the dependence on annotated information from a few positive samples. The cross graph attention mechanism used can also visualize the most significant molecular substructures or protein residues that contribute to prediction, providing valuable intermediate level explanations for the model's decision-making.

3 Prediction of drug target interactions based on protein language model

Protein language models pre-trained on large-scale protein sequence libraries based on Transformer architecture, such as Evolutionary Scale Model-2 (ESM-2), have encoded billions of years of evolutionary information into powerful context aware vectors. The current research is committed to going beyond the initial stage of using it as a static feature input and exploring deeper ways of model integration.

Graph Positional encoding and Sequence features for Drug-Target Interaction Prediction embodies a collaborative paradigm where protein language models are responsible for feature extraction and graph neural networks are responsible for relational inference [4]. This framework uses graph isomorphism networks to encode drug molecule graphs and ESM-2 to encode protein sequences. Its core lies in a multi head cross attention module that allows for bidirectional, iterative attention interaction between drug and protein features in multiple representation subspaces. This process to some extent simulates the dynamic mechanism of induced binding in molecular recognition, where the binding pocket conformation of proteins can be adaptively adjusted due to the presence of ligands, resulting in a joint feature representation that better reflects real biophysical processes.

The potential of protein language models goes beyond providing sequence features. Studies have shown that ESM-2's self-attention map can accurately predict spatial contacts between protein residues, which is closely related to AlphaFold2's ability to predict protein structure [5]. Recent studies such as i-Fold have attempted to integrate the residue importance scores predicted by language models into the geometric folding module of AlphaFold2, in order to further improve the accuracy of 3D structure prediction [6].

4 Prediction of drug target interactions based on geometric deep learning

The binding between drugs and targets ultimately occurs in three-dimensional physical space, and their affinity is highly dependent on the stereo complementarity of the two parties in terms of shape, electrostatics, hydrogen bonding, and other aspects. Geometric deep learning, through the introduction of equivariant neural networks, enables models to strictly follow the symmetry of rotation, translation, and other aspects of three-dimensional space, making it an ideal tool for processing 3D molecular and protein structures [7].

The design of equivariant graph neural networks ensures that their outputs undergo corresponding collaborative transformations with the rotation or translation of input coordinates. In DTI prediction, when inputting the atomic 3D coordinates of drug molecules and protein binding pockets, each layer of the equivariant graph neural network (EGNN) will

simultaneously update the coordinate vector and feature vector of each atom: the update of coordinates depends on the current feature information, and the update of features depends on the relative position and direction between atoms. This enables the model to synchronously optimize the conformation of molecules during the learning process, and the learned feature representations include stereochemistry and steric hindrance effects. Applying such models to process protein 3D structures predicted by AlphaFold2 can more realistically simulate spatial matching and changes in interaction energy during drug target binding.

Geometric deep learning models have been proven to outperform traditional graph neural networks in predicting the quantum chemical properties of molecular systems. This indicates that combining quantum mechanics calculations with geometric deep learning to introduce deeper physical and chemical principles for DTI prediction is expected to solve the interpretability problem of current models that mainly rely on experimental data statistical fitting and lack a solid physical foundation.

5 Drug target interaction prediction based on generative artificial intelligence

Generative artificial intelligence is an artificial intelligence technology that can learn the underlying distribution from data and generate new and reasonable data samples. In the field of DTI, generative artificial intelligence, represented by diffusion models, is transforming from auxiliary tools to one of the core engines driving research innovation, with its impact running through the entire process of data, models, and molecular design.

In terms of data augmentation and conformational generation, diffusion models have brought breakthroughs. The traditional generation of molecular conformations relies on computationally expensive molecular dynamics simulations. The equivariant consistency model (EC-Conf), based on the principle of equivariant diffusion, can generate a molecular three-dimensional structure ensemble that conforms to physical laws and covers conformational space at an extremely fast speed. This is crucial for DTI prediction, as the active conformation of a drug may not necessarily be its most stable, lowest energy conformation. Providing diverse and reasonable conformations as training data for the model can significantly enhance its ability to model molecular flexibility and conformational changes [8]. In addition, the graph diffusion model can directly operate in the representation space of the molecular graph, generating novel molecular structures with specific properties through denoising processes, and combining them with DTI prediction models to form an active design closed loop from generating candidate molecules to evaluating binding potential [9].

A more promising direction lies in the integration of Large Language Model (LLM) and diffusion models. A system that integrates LLM and 3D diffusion models can parse text instructions, generate 3D molecular structures that meet the requirements, and evaluate their binding strength through DTI prediction modules. This end-to-end capability, from natural language intent to three-dimensional molecular generation and functional evaluation, is expected to completely reshape the development model of early drug discovery.

6 Challenge and future

6.1 Core bottleneck

The contradiction between model complexity and computational efficiency is becoming increasingly acute. The integration of 3D structures, large language models, and diffusion generation components makes the most advanced models exceptionally large, and the cost of single inference or generation is high, making it difficult to directly apply to industrial scenes that require rapid virtual screening of tens of millions of compound libraries. Developing model compression, knowledge distillation, and efficient attention algorithms has become an equally urgent task as model innovation.

The level of interpretability of the model is still relatively shallow. Existing methods often identify atoms or residues that contribute significantly to prediction, but fail to provide a deep understanding of the biomechanical mechanisms underlying their interactions and how they affect downstream signaling pathways or disease phenotypes. To achieve deeper mechanistic explanations, it may be necessary to deeply integrate and infer deep learning models with biological knowledge graphs that contain causal relationships at the symbolic level.

The bias and noise of the data itself constrain the generalization ability of the model. The public dataset is highly concentrated on certain target families with in-depth research, resulting in unstable performance of the model on obscure or completely new categories of targets [10]. Meanwhile, the combined data obtained from the experiment itself contains false positives and false negatives, which pose inherent difficulties for model learning due to annotation noise.

6.2 Evolution Path

The future development direction will be to build a unified multimodal pre-trained large model of biomolecules. Such an ideal system framework may have the following features: its input interface can be compatible with Simplified Molecular Input Line Entry System (SMILES) strings, molecular graphs, 3D coordinates, protein sequences, structures, and even scientific literature abstracts and other multimodal data; Its core adopts a layered architecture, with the bottom layer consisting of dedicated encoders that handle different modalities, the middle layer being a universal module for cross modal alignment and fusion, and the upper layer being a pluggable prediction head for specific tasks such as DTI prediction and toxicity assessment; Its training first involves self supervised pre training on massive cross modal biological data to learn

universal molecular representations and biological laws, and then quickly adapts to downstream tasks through parameter efficient fine-tuning techniques; Its output is not only predicting scores, but also generating hypotheses that combine patterns, proposing molecular optimization suggestions, and displaying them through an interactive interface.

7 Conclusion

The field of DTI prediction is undergoing a critical transformation from passive prediction to active design. The deep integration of heterogeneous graph neural networks, protein language models, and geometric deep learning provides a solid technical foundation for overcoming the challenges of data and knowledge representation. The empowerment of generative artificial intelligence further expands the exploration scope from known chemical spaces to broader unknown fields.

There are still several core bottlenecks in the deployment of advanced models to practical applications, and the contradiction between model complexity and computational efficiency is becoming increasingly prominent. Although integrating multimodal and large parameter models improves performance, their high computational costs constrain the practicality of rapid drug screening. Although these models can identify key atoms or residues, they are difficult to delve into the biological mechanisms behind their interactions. Meanwhile, the performance of the model is also limited by the quality and coverage of the training data, which affects its generalization ability and the credibility of the results.

The key to realizing the vision of intelligent drug discovery lies in building the next generation of AI systems that can uniformly understand multimodal biological information, have efficient computing capabilities, and decision-making processes that can be understood and trusted by biologists. This requires deeper interdisciplinary collaboration between computer science, computational chemistry, and experimental biology.

References

1. M. Gao, D. Zhang, Y. Chen, et al., GraphormerDTI: a graph transformer-based approach for drug-target interaction prediction. Comput. Biol. Med. 173, 108339 (2024)
2. M. Wang, X. Lei, L. Guo, et al., DHGT-DTI: advancing drug-target interaction prediction through a dual-view heterogeneous network with GraphSAGE and graph transformer. J. Pharm. Anal. 101336 (2025)
3. W. Dai, Y. Wang, S. Yan, et al., Heterogeneous network drug-target interaction prediction model based on graph wavelet transform and multi-level contrastive learning. Sci. Rep. 15(1), 30326 (2025)
4. A. Xiong, Z. Luo, Y. Xia, et al., An interpretable geometric graph neural network for enhancing the generalizability of drug-target interaction prediction. BMC Biol. 23(1), 350 (2025)
5. Y. Liu, B. Kang, Q. Cui, Enhancing protein structure prediction accuracy by prioritizing important residues using protein language models. bioRxiv 2025.09.28.679101 (2025)
6. T.T. Nguyen, Z. Jiang, V.N. Nguyen, et al., Integrating ESM-2 and graph neural networks with AlphaFold-2 structures for enhanced protein function prediction. ACS Omega 10(33), 38103-38111 (2025)
7. J. Sun, Y. Cao, H. Hu, et al., Equivariant learning leveraging geometric invariances in 3D molecular conformers for accurate prediction of quantum chemical properties. Sci. Rep. 15(1), 26969 (2025)
8. X. Chen, Y. Shen, M. Mahmud, et al., ConnectomeDiffuser: Generative AI enables brain network construction from diffusion tensor imaging. IEEE Trans. Consum. Electron. (2025)
9. B. Benjdira, A.M. Ali, W. Boulila, et al., Fusing LLMs and diffusion models: A comprehensive survey of progress, challenges, and future directions in generative AI. Comput. Sci. Rev. 60, 100881 (2026)
10. X. Wang, Q. Gao, H. Zhang, et al., MAAPE: a tool for modular evolution analysis of protein embeddings. Digit. Discov. 4(11), 3245-3259 (2025)