Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Overview of AI-Driven New Drug Development Skills

Xiongpeng Jiang

School of Mathematical Sciences, Zhejiang University of Technology, Hangzhou, Zhejiang 310000, China

| ABSTRACT

In traditional drug development, long development cycles, high costs, and low efficiency are often unavoidable. With the rise of artificial intelligence (AI), all aspects of traditional pharmaceutical industry have undergone unprecedented changes. Through model training, drug development cycles can be significantly shortened and efficiency improved. This article reviews the changes brought about by AI to the drug development industry in the past three years, outlining the improvements AI has made to the pharmaceutical process, and contributing to future research in the pharmaceutical field.

| KEYWORDS

Artificial intelligence; drug discovery; machine learning; deep learning; vaccine development

| ARTICLE INFORMATION

ACCEPTED: 20 March 2026 **PUBLISHED:** 09 April 2026 **DOI:** 10.32996/jcsts.2026.8.5.12

1. Introduction

Drugs are an indispensable part of our daily lives. Usually, it takes 10 to 15 years for a new drug to go from research and development to market launch. According to estimates, the cost of developing a new drug has now reached $3 billion. Despite the large amount of investment and effort, the success rate of drug development is still very low. The overall success rate in the clinical stage is only 10.4% [1]. Drug development is trapped in a dilemma of long cycle, high technology, and high risk. The low success rate and high investment of traditional drug development methods are no longer sustainable.

With the rapid development of artificial intelligence (AI) technology, the drug development process has also achieved unprecedented development. The concept of "artificial intelligence" appeared around 1950s, and machine learning (ML) appeared around 1980. The research paradigm of AI has moved from artificial rules to data mining. ML models can learn autonomously from data and summarize the rules. Traditional ML algorithms include decision trees, random forests, K-nearest neighbors, support vector machines, etc. [2]. With the breakthrough of deep learning technology, AI has shown strong capabilities in solving key professional problems. High-precision prediction models have been used to evaluate the bioactivity and safety of different drug forms [3]. The multi-feature fusion algorithm RCI (Redundancy-Correlation and Interaction) based on the combination of redundancy-correlation and interaction analysis has shown high accuracy in building drug target prediction models and improved the representation ability of drug target fusion data in the model.[4]

This article reviews and introduces the AI-driven drug development process in recent years, and proposes the challenges facing AI in the field of drug development and future research directions.

2. AI applied to drug-target interactions

Drug targets refer to molecules in the body that are intrinsically linked to specific disease processes and can bind with drugs to produce therapeutic effects. In the process of new drug development, one of the first problems to be solved is the identification of drug-target interaction (DTI), that is, determining whether drug molecules and targets will interact [1], and using this process to

find drug molecules that interact with a specific target. Compared with traditional calculation methods, using AI to help predict DTI can accelerate the drug development process and reduce the development cost.

2.1 Drug-target interaction prediction methods

2.1.1Ligand-based prediction methods

Ligand-based prediction methods use known ligand information of the target to predict the interaction between the new ligand and the target, and are used to predict the bioactivity of the molecule on a specific target [1]. The most commonly used ligand-based prediction method is the quantitative structure-activity relationship (QSAR) method [5]. This method uses the structure of the compound to build a model to analyze the interaction between the molecule and the target. In detail, this method evaluates the interaction between the molecule and the target by comparing the similarity between the structure and activity, that is, by comparing the similarity between the molecule and the known ligand structure [6]. The QSAR method can screen molecules without knowing the target structure, but it has the lack of clear physical interpretability and cannot well understand the mechanism of action.

2.1.2 Structure-based prediction methods

Structure-based prediction methods predict the interaction between drug molecules and target proteins based on their spatial structures. One classic method is molecular docking. This method studies the affinity between drug molecules and target proteins based on the principle of structural and energetic complementarity [7]. Common molecular docking methods can be classified as flexible docking, semi-flexible docking, and rigid docking [1]. Molecular docking is fast and accurate, but it cannot be used without structural information [8].

2.1.3 Prediction methods based on chemical genomics

This method extracts the biological characteristics of drug molecules and target proteins and predict their interaction based on a predictive model. There are generally three approaches to this method: ligand learning based on target families, inference of target common ligands based on ligand binding sites, and prediction of drug-target relationship based on ligand-receptor interaction [9]. The advantage of this type of method is that there is a large amount of data available, but the disadvantage is that when the interaction between the "drug-target pair" is not confirmed, its predictive performance will be limited [10]. However, we can also see from this that the prediction method based on chemical genomics is very suitable for combination with AI models, because the amount of pharmacological spatial data is huge, and AI can quickly and effectively extract information from massive amounts of data [1].

2.1.4 Application of Artificial Intelligence in Drug-Target Interaction Prediction

AI can predict interactions and target proteins based on existing information about the interaction between drug molecules and target proteins, thereby quickly and efficiently identifying candidate drugs for subsequent experiments. In general, AI methods applied to DTI prediction can be divided into four categories: similarity-based methods, feature-based methods, network-based methods, and deep learning-based methods [1].

Similarity-based methods are mainly based on the relationship between similarity and interaction. This method scores the similarity of targets or drug molecules and then predicts DTI based on the score results[1]. Shi et al. developed a nonparametric, memory-based method, SRP (similarity rank-based predictor), which uses similarity and similarity ranking to calculate two indices—trend index and counter-trend index—to represent the probability of drug molecules interacting with target proteins and the probability of not interacting, respectively. It achieved high accuracy on benchmark datasets [1]. Feature-based methods encode the biological information of drug molecules and target proteins, generate feature descriptors to describe molecules and targets, integrate these descriptors into feature vectors, and then apply AI algorithms for prediction [1]. For example, Cao et al. used the random forest method, combined with chemical, biological and network feature calculation methods to predict DTI, and applied it to the four types of DTI networks (enzymes, ion channels, G protein-coupled receptors and nuclear receptors) in the human body to verify the model performance [1]. Network-based methods use complex network graphs to mine the connection between drug molecules and target proteins, thereby achieving DTI prediction. For example, Olayan et al. developed a model called DDR, which enables... Multiple similarity indicators were integrated using a nonlinear fusion method to construct a heterogeneous network of drug targets, and the random forest method was used to find drug target interaction pairs in the heterogeneous network [1]. The deep learning-based method constructs feature vectors by using biological signals of drug molecules and target proteins, and then inputs them into a deep neural network for training to predict DTI. For example, Pen et al. proposed a method based on representation learning and deep neural networks. This method first uses Jaccard similarity coefficient and random walk with restart (RWR) to extract relevant features of drugs and proteins from the heterogeneous network, then uses a denoising autoencoder to reduce the dimension, and constructs a CNN to predict DTI [1].

In summary, the application of AI to drug-target interaction (DTI) prediction can accelerate the speed of DTI prediction, significantly shorten the drug development cycle, and, thanks to AI's ability to quickly sift information from massive amounts of data, we can perform high-throughput screening of virtual compound libraries, greatly expanding the scope of prediction and significantly reducing the cost and number of experiments required by traditional experimental screening. However, AI-assisted prediction also faces challenges such as high data dependence and poor interpretability. In the future, we should improve the interpretability of AI and avoid problems such as overfitting and poor generalization.

3. AI applied to molecular generation

Unlike traditional virtual screening, molecular generation and optimization have gradually become important means of drug screening. This approach can directly generate new molecules with specific properties in chemical space, realizing the integration of "design-screening" and accelerating the discovery of active compounds [11]. For example, the SE(3)-equivariant diffusion model has achieved excellent performance in the three-dimensional generation and flexible docking of protein-ligand complexes due to its invariance under rotation and translation of three-dimensional molecular structures [12].

Compared with traditional molecular generation, AI-enabled molecular generation has higher efficiency and scalability. Through multi-objective optimization and direct target-oriented optimization [13], experimental costs and resource consumption can be greatly reduced. However, AI-enabled molecular generation also faces problems such as heavy reliance on data and high field threshold. Future experimental research should focus on improving model interpretability and solving problems such as poor data quality.

4. AI applied to virtual screening

Unlike rule- or structure-driven limited searches, AI-assisted virtual screening is a data- or model-driven intelligent prediction and large-scale exploration. Virtual screening refers to the rapid prediction and screening of molecules with potential binding activity to the target in a compound library[14], thereby narrowing the experimental scope and reducing experimental costs.

Compared with traditional virtual screening, AI-driven virtual screening has shown significant advantages in the development of modern drug discovery due to its data-driven, high-throughput computing and autonomous feature mining capabilities. In response to the limitations of traditional virtual screening in terms of throughput, generalization ability and novel/unexplored targets, Lan et al. [15] proposed the DrugCLIP model in 2026. DrugCLIP extracts high-dimensional features through protein-molecule dual encoders and achieves cross-modal feature alignment by using contrastive learning. Through these methods, it has completed ultra-high-throughput AI virtual screening at the whole genome scale, breaking through the speed and scale limitations of traditional screening. Although AI-driven virtual screening has significant advantages, it still faces many challenges in actual use, such as high-quality data and interpretability (black box problem). Future experimental research should pay more attention to the interpretability and data dependence of the model.

5. AI applied to vaccine development

Vaccine research and development is a core research direction in the public health system and clinical medicine. Vaccines are not only a fundamental means of preventing and controlling infectious diseases, but also play an irreplaceable role in public health security, social and economic development, and medical progress. Traditional vaccine research and development based on classical immunology and biological experimental systems is technically mature, but it has problems such as long research and development cycle, difficulty in dealing with sudden infectious diseases, low research and development efficiency, and high research and development cost. With the rapid development of AI, AI models based on machine learning algorithms can learn from massive amounts of data, which greatly shortens the research and development cycle, greatly reduces the research and development cost, improves the research and development accuracy, and can also deal with sudden infectious diseases. For example, Hou Yanjun et al.[16] developed a messenger RNA bidirectional encoder representation from Transformers (mRNABERT), which is a messenger RNA (mRNA) language model based on the architecture of bidirectional encoder representation from Transformers (BERT). The model can learn from massive amounts of mRNA sequence data, grasp its structural and functional rules, accelerate the sequence design of mRNA vaccines, and also improve stability. In addition, Tang et al.[17] integrated AI technology into the development of therapeutic vaccines for pancreatic cancer. By targeting neoadjuvant chemotherapy (NACT)-induced tumor metabolic reprogramming, they regulated the immune landscape of the tumor microenvironment (TME). This approach significantly enhanced the anti-tumor immune response and synergistic effect of the therapeutic vaccine with chemotherapy.

6. Summary

Artificial intelligence (AI) has been widely applied throughout the entire drug development process (target discovery, compound design and screening, preclinical research, clinical trials, and drug production and quality control), significantly improving efficiency, success rate, and cost. AI's ability to process massive amounts of data enables it to accurately predict drug targets, optimize drug molecular structures, and shorten development cycles and reduce costs. Furthermore, in vaccine development, AI can efficiently perform antigen screening, sequence design, and immunogenicity assessment, preparing for emerging infectious diseases.

While AI has brought numerous benefits to drug development, it still has some shortcomings. For example, AI model training relies on high-quality datasets; insufficient data quality or quantity can lead to biased predictions. Furthermore, AI models often lack interpretability, making it difficult to accurately explain the mechanisms of drug development. Future research will focus on building standardized, high-quality dataset platforms to improve the interpretability and reliability of AI models.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Li, Q. Y., Zhang, X. C., & Wang, S. Q. (2022). Research progress of artificial intelligence in predicting drug-target interactions. Chinese Journal of Pharmacology and Toxicology, 36(1), 1-10. (in Chinese)
[2] Tang, Q., & Dou, J. (2026). Research progress on the application of artificial intelligence in vaccine research and development. Chinese Journal of Immunology, 42(1), 247-256. (in Chinese)
[3] Zhao, D. H., & Xie, C. Y. (n.d.). Application progress of artificial intelligence agents in drug discovery and development. Journal of Zhejiang University (Medical Sciences), 1-12. (in Chinese)
[4] Zhang, X. Y., Chen, Z., Huang, Y., et al. (2024). Drug-target interaction prediction based on multimodal stacked hybrid autoencoder. Computer Engineering and Applications, 60(19), 334-342. (in Chinese)
[5] Dudek, A. Z., Arodz, T., & Gálvez, J. (2006). Computational methods in developing quantitative structure-activity relationships (QSAR): A review. Combinatorial Chemistry & High Throughput Screening, 9(3), 213-228.
[6] Speck-Planche, A., & Cordeiro, M. N. D. S. (2014). QSAR multi-target in drug discovery: A review. Current Computer-Aided Drug Design, 10(2), 129-136.
[7] Pujadas, G., Vaqué, M., Ardèvol, A., Bladé, C., Salvadó, M. J., & Arola, L. (2008). Protein-ligand docking: A review of recent advances and future perspectives. Current Pharmaceutical Analysis, 4(1), 1-19.
[8] Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W., & Kanehisa, M. (2008). Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. Bioinformatics, 24(13), i232-i240.
[9] Jacob, L., & Vert, J. P. (2008). Protein-ligand interaction prediction: An improved chemogenomics approach. Bioinformatics, 24(19), 2149-2156.
[10] Chen, R., Liu, X., Jin, S., Lin, J., & Liu, J. (2018). Machine learning for drug-target interaction prediction. Molecules, 23(9), 2208.
[11] Cao, W. X., Zhang, X., Yang, H., et al. (2026). Research progress on the application of artificial intelligence in high-throughput drug screening. Acta Pharmaceutica Sinica, 61(2), 356-364. (in Chinese)
[12] Zhu, J., Gu, Z., Pei, J., Lai, L., & Shao, B. (2024). DiffBindFR: An SE(3) equivariant network for flexible protein-ligand docking. Chemical Science, 15(21), 7926-7942.
[13] Elton, D. C., Boukouvalas, Z., Fuge, M. D., & Chung, P. W. (2019). Deep learning for molecular design—A review of the state of the art. Molecular Systems Design & Engineering, 4(4), 828-849.
[14] Lavecchia, A., & Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: A critical review. Current Medicinal Chemistry, 20(23), 2839-2860.
[15] Jia, Y., Gao, B., Tan, J., et al. (2026). Deep contrastive learning enables genome-wide virtual screening. Science, 391(6781), eads9530.
[16] Xiong, Y., Wang, A., Kang, Y., et al. (2025). mRNABERT: Advancing mRNA sequence design with a universal language model and comprehensive dataset. Nature Communications, 16(1), 10371.
[17] Tang, R., Xu, J., Wang, W., et al. (2023). Targeting neoadjuvant chemotherapy-induced metabolic reprogramming in pancreatic cancer promotes anti-tumor immunity and chemo-response. Cell Reports Medicine, 4(10), 101234.