The role of artificial intelligence in drug screening, drug design, and clinical trials

Yuyuan Wu[1†], Lijing Ma[1†], Xinyi Li[1], Jingpeng Yang[1], Xinyu Rao[1], Yiru Hu[1], Jingyi Xi[1,2], Lin Tao[1,2], Jianjun Wang[3], Lailing Du[4*], Gongxing Chen[1,2*] and Shuiping Liu[1,2*]

[1]School of Pharmacy, Hangzhou Normal University, Hangzhou, Zhejiang, China, [2]Key Laboratory of Elemene Class Anti-Cancer Chinese Medicines, Engineering Laboratory of Development and Application of Traditional Chinese Medicines, Collaborative Innovation Center of Traditional Chinese Medicines of Zhejiang Province, Hangzhou Normal University, Hangzhou, Zhejiang, China, [3]Department of Respiratory Medicine of Affiliated Hospital, Hangzhou Normal University, Hangzhou, Zhejiang, China, [4]Key Laboratory of Pollution Exposure and Health Intervention of Zhejiang Province, Shulan International Medical College, Zhejiang Shuren University, Hangzhou, China

The role of computational tools in drug discovery and development is becoming increasingly important due to the rapid development of computing power and advancements in computational chemistry and biology, improving research efficiency and reducing the costs and potential risks of preclinical and clinical trials. Machine learning, especially deep learning, a subfield of artificial intelligence (AI), has demonstrated significant advantages in drug discovery and development, including high-throughput and virtual screening, *ab initio* design of drug molecules, and solving difficult organic syntheses. This review summarizes AI technologies used in drug discovery and development, including their roles in drug screening, design, and solving the challenges of clinical trials. Finally, it discusses the challenges of drug discovery and development based on AI technologies, as well as potential future directions.

1 Introduction

New drug development includes the discovery of drug lead compounds and drug optimization, which is a long, expensive, and high-risk process. The transformation of a new drug from a promising candidate to a marketable product can take over a decade or more, cost up to a billion dollars, and result in a high rate of clinical failure (Liu C. et al., 2021). Currently, diseases such as cancer, diabetes, Alzheimer's disease, and Parkinson's disease significantly affect human health and have become a serious public health problem globally, making drug discovery and development increasingly critical (Yuan et al., 2023; Zheng and Jiang, 2019). Drug development companies have adopted various methods to overcome this dilemma, with artificial intelligence (AI) playing a key role. For example, a study by the technology company Tech Emergence revealed that applying AI to new drug development can speed up the process by 2%, and a report by Goldman Sachs predicted that as AI technology matures, the annual savings in the field of new drug development could be as high as 28 billion dollars (Mao and Liu, 2021). The number of innovative drugs approved in China's drug market was 14, 24, 18, 37, 49, 48, and 47 between 2014 and 2020, with the number of

domestically produced innovative drugs increasing from 0 in 2017 to 14 in 2020 (Tang et al., 2022). These figures are attributed to the undeniable influence of AI on this process, indicating that AI is anticipated to revolutionize drug development.

2 Technologies and algorithms related to AI in drug discovery and development

The concept of AI dates back to 1950, when scientist Alan Turing described a simple test, later known as the "Turing Test" in his book Computing Machinery and Intelligence, to determine whether a computer exhibited human intelligence. He described AI as similar to but more complex than the human brain. Turing is thus known as the "Father of AI" (Mintz and Brodie, 2019; Kaul et al., 2020).

Machine learning (ML) is a subfield of AI, with deep learning (DL) as a subset of ML (Pandiyan and Wang, 2022). Currently, AI can analyze more complex algorithms and perform DL. Many related algorithmic models have been developed for drug discovery. ML algorithms have been used in several drug discovery processes, including peptide synthesis, structure-based virtual screening, ligand-based virtual screening, toxicity prediction, drug monitoring and release, pharmacodynamic modeling, quantitative structure-activity relationships, drug repositioning, polypharmacology, and physicochemical activities (Gupta et al., 2021). The relevant algorithmic models based on ML are described below.

2.1 ML

ML refers to AI algorithms in which models are trained on large datasets to learn rules, analyze new data, and make predictions and decisions. There are three main types of ML: supervised learning, unsupervised learning, and reinforcement learning (Figure 1) (Pandiyan and Wang, 2022; Tuan et al., 2023; Mak and Pichika, 2019). Supervised learning involves training algorithms on labeled datasets with predetermined correct answers for each input, enabling accurate predictions of new, unseen inputs (Islam et al., 2024). Unsupervised learning recognizes hidden patterns in data, clusters them, and interprets them in groups, with outputs such as disease subtypes and target discovery (Mak and Pichika, 2019). Reinforcement learning entails learning from interactions with the environment (Islam et al., 2024). For instance, Chen et al. used the Trinity software to assemble 148,784 transcripts and 78,092 single genes from clean reads. Expression patterns or functionally relevant gene clusters could be identified under specific conditions by analyzing the assembled gene expression data using the ML approach (Chen et al., 2021).

DL, a ML technique based on various types of neural networks (NNs) that utilize a hierarchical structure to learn more complex structures and relationships in a dataset (Smith et al., 2023), is a type of modeling and learning that employs neuronal structures mimicking biological neural networks (Figure 1). DL differs from ML because it utilizes data types and learning methods more adept at handling massive, high-dimensional, and complex data structures (Lac et al., 2024). DL eliminates some of the data preprocessing steps associated with ML and continuously corrects prediction errors
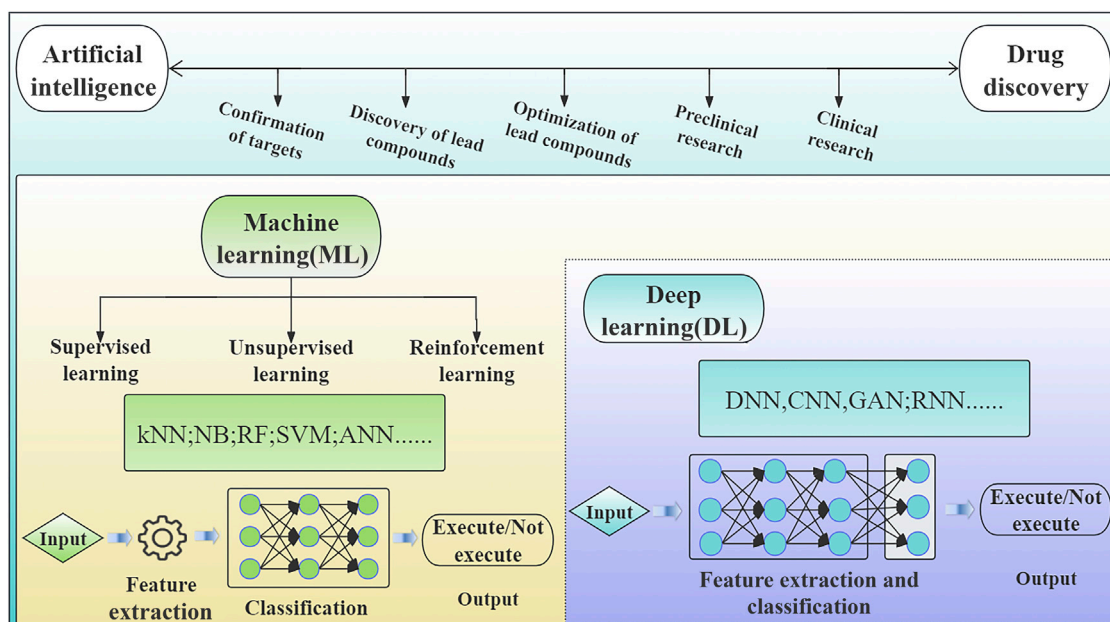
FIGURE 1
An overview of AI applications and technologies in drug discovery. The use of AI in drug discovery encompasses various applications and techniques The integration of AI into drug discovery involves the incorporation of ML and DL, with DL being a subset of ML. ML is further classified into three primary types: supervised, unsupervised, and reinforcement learning. This figure also delineates the ML and DL algorithms, highlighting the differences in their execution methodologies.

through gradient descent and backpropagation. DL techniques have demonstrated significant potential and prospects for various clinical applications, including medical image analysis, disease diagnosis, treatment prediction, and patient monitoring (Wang et al., 2024).

2.2 Traditional ML algorithms

The most common ML algorithms in drug discovery and development include k-Nearest Neighbors (kNN), Naïve Bayesian Classifier (NB), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Networks (ANNs) (Figure 1). Their roles in drug discovery and development have been summarized as following.

2.2.1 kNN

kNN implies that a sample belongs to a specific category if the majority of the k-nearest samples (the closest neighbors in the feature space) surrounding it belong to that category (Guan et al., 2023). In recent studies, Yang M et al. used the weighted kNN (WkNN) method to improve the overall density of the drug-disease association matrix based on the kNN principle for drug repositioning research (Yang et al., 2023).

2.2.2 NB

NB is one of the Bayesian classifiers that can be used to train a model with a dataset of known categories, enabling the categorization of data from unknown categories (Yang et al., 2023). NB has been used in the pharmaceutical field due to its simplicity, speed, and effectiveness. For instance, based on the principle of NB, Shi H et al. trained a classifier to recognize positive and negative samples of the pregnane X receptor (PXR). This classifier was then used to distinguish between PXR activators and non-activators, thereby improving classification efficiency (Shi et al., 2015).

2.2.3 RF

The RF is a regression tree technique that uses bootstrap aggregation and randomization of predictor variables to achieve a high degree of predictive accuracy (Rigatti, 2017). Recently, Ryu et al. (2022) developed a computational model called PredMS based on the RF model to predict the stability of small compounds during metabolism.

2.2.4 SVM

The SVM is a two-class classification model. Its basic model is defined as a linear classifier that maximizes the intervals on the feature space. The learning strategy of SVM is interval maximization, which ultimately translates to solving a convex quadratic programming problem (Yang et al., 2023). It is crucial for predicting molecular interactions, binding affinity, and other properties between ligands and target proteins (Shiammala et al., 2023). Jing-Fang Z et al. selected 324 neurotoxic compounds and 234 non-neurotoxic compounds based on the combination of SVM with Cfs subset evaluation and Best First-D1-N5 search using a web database. Compounds were used as the dataset for constructing the neurotoxicity discriminant model. This dataset included charge distribution and physicochemical and geometric descriptors to

characterize the molecular structures of neurotoxic compounds. The accuracy, sensitivity, and specificity were >80% (Zhang et al., 2014).

2.2.5 ANNs

ANNs are computer programs that simulate the operation of many processing units that mimic nerve cells and the basic biological mechanisms by which they connect and interact with each other. ANNs are a subset of ML that were created as direct analogs of biological NNs (Shiammala et al., 2023). Like the human brain, ANNs can learn from experiences and understand the general relationships between variables. These algorithms are crucial for various processes, including drug screening and design.

Despite the unique characteristics of each of these ML-based algorithmic models, they also have limitations. First, these models often do not consider the heterogeneous information defined in relational networks. Second, AI/ML-based models require extensive training, and each application necessitates specific training tailored to its requirements. Additionally, shallow network- and sequential data-based approaches are usually insufficient to learn some of the key features (e.g., distance correlation) required to make accurate predictions (Wu et al., 2024).
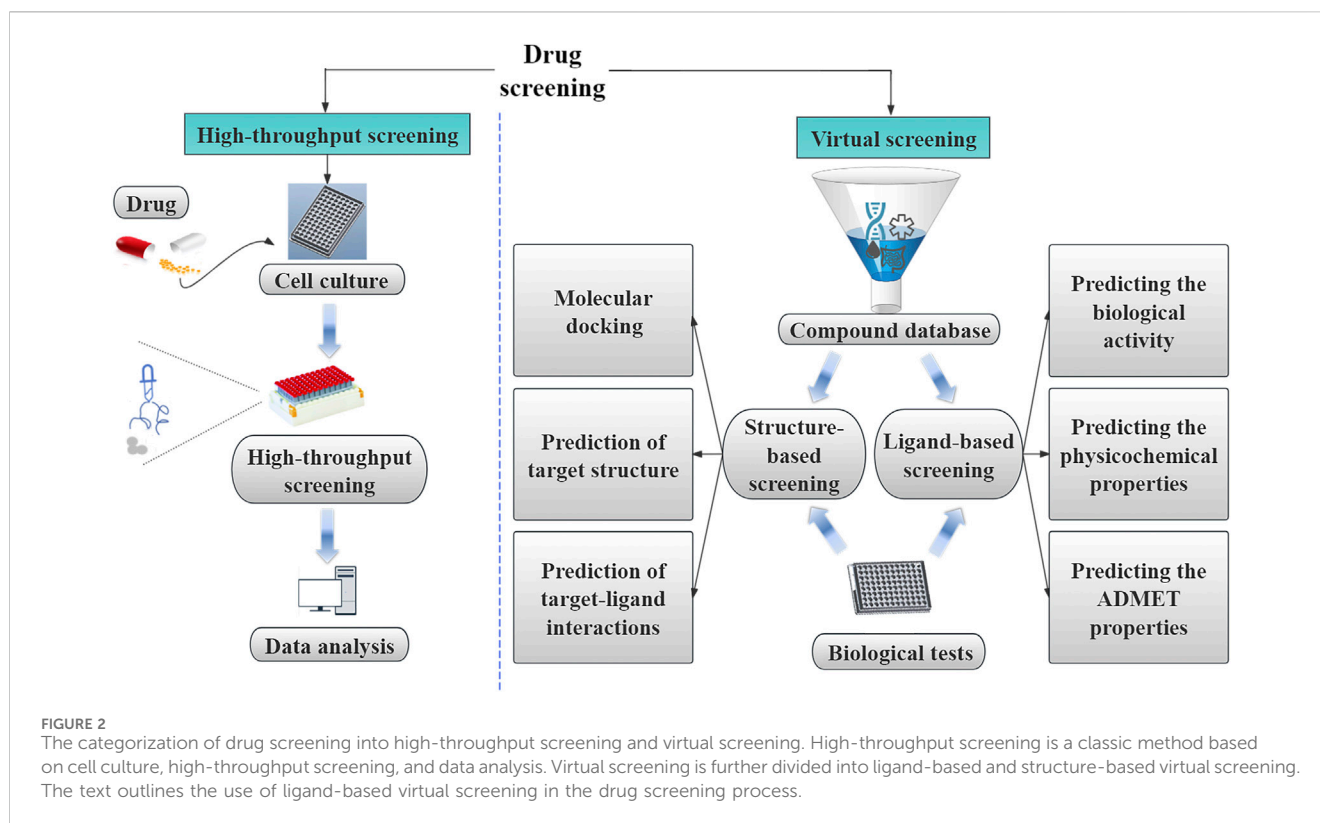
2.3 DL-based algorithm

DL algorithms for drug discovery usually consist of convolutional neural networks (CNNs), generative adversarial networks (GANs), and recurrent neural networks (RNNs) (Figure 1). All of them play critical role in drug discovery and development, which have been summarized as following.

2.3.1 CNNs

CNNs operate with a convolutional layer that slides over the original image using convolutional filters (typically small matrices of 3×3 or 5×5 in size), allowing each filter to extract specific features, thereby reducing the amount of computation and the risk of overfitting after maximum pooling and average pooling by a pooling layer. These are compressed into a lengthwise vector that serves as an input to a fully connected layer. The fully connected layer then uses these features to determine image categories. Lei et al. (2021) developed a learning framework based on DL called CAMP, which utilizes CNNs and self-attention mechanisms to adequately extract local and global information to predict binary interactions of input peptide-protein pairs.

2.3.2 GANs

The operation of a GAN involves a generator and a discriminator. Random inputs are passed through the generator to produce new samples, which are then given to the discriminator to distinguish between real and fake. These two components continuously challenge each other to generate more authentic sample data (Gulakala et al., 2022). For instance, Wang et al. (2023) constructed a new CNN using dense networks. Dense networks perform multilayer transmission on the generator network of the GAN architecture, extending the training space and improving sequence generation efficiency (Wang et al., 2023).

FIGURE 2
The categorization of drug screening into high-throughput screening and virtual screening. High-throughput screening is a classic method based on cell culture, high-throughput screening, and data analysis. Virtual screening is further divided into ligand-based and structure-based virtual screening. The text outlines the use of ligand-based virtual screening in the drug screening process.

2.3.3 RNNs

RNNs are particularly important for analyzing information based on sequences or time series (Pandiyan and Wang, 2022), which are distinguished by their ability to process image and numerical data and learn data types exhibiting forward and backward correlations due to the network's inherent ability to memorize them. Based on the advantages of RNNs for data processing, Sangrak et al. constructed an RNN model that significantly improved the performance of drug interaction extraction by incorporating positional features, subtree inclusion features, and integration methods. Compared to the top models in the same period, this model exhibited improved performance on the DDIExtraction Challenge'13 test data by up to 4.4% and 2.8%. This improvement provides an effective solution for large data processing and plays a key role in drug discovery (Lim et al., 2018).

Despite its widespread use, AI-based DL requires special conditions. It starts with the need for high-quality and sufficiently large data, which are often privately owned and not intuitively generated. Additionally, data in ML-generated "black boxes" can be difficult to interpret, especially in the fields of biology and chemistry (Lei et al., 2021).

3 AI opens a new chapter in drug screening

Drug screening involves the identification and evaluation of the initial pharmacological properties of substances with potential medicinal applications aming to uncover their therapeutic value and clinical utility. This process is a foundational step in the research

and development of novel drugs, providing essential data and insights. Drug screening is commonly divided into two main categories: high-throughput and virtual screening (Figure 2).

High-throughput screening uses experimental techniques at the cellular or molecular level, as well as microtiter plates, automated systems, and rapid detection instruments for data collection, analysis, and processing (Figure 2). This approach involves screening thousands or even millions of compounds during the drug discovery phase, serving as a crucial technique for identifying active compounds early in the drug development process (Figure 2) (Dueñas et al., 2023). For instance, in the context of ATR kinase inhibitor research, high-throughput screening using compound libraries has proven to be an effective strategy for discovering ATR kinase inhibitors, exemplified by the identification of the lead compound BAY-937 (Yuan et al., 2022). In antiviral research targeting articulator-associated protein kinase 1 (AAK1), Qi et al. used high-throughput RNAi screening to determine the pivotal role of AAK1 in regulating the entry of rabies virus into cells (Qi et al., 2022). Furthermore, in a study investigating the relationship between intestinal flora and Shengmai Yin, You et al. used quantitative sequencing technology to assess alterations in the composition of the intestinal microbial community in a rat model of splenic deficiency (You et al., 2020).

Virtual screening plays a crucial role in identifying potential drug candidates by evaluating and screening extensive structural libraries (Figure 2) (Singh et al., 2024). Using software tools, virtual screening enables the simulation of molecular interactions, calculation of affinities, and streamlining screening processes to improve efficiency. Virtual screening is particularly advantageous for drug discovery involving large numbers of small molecules, with

libraries exceeding 10^{60} compounds within the total chemical space, surpassing the capacity of traditional high-throughput screening methods that are limited to tens to hundreds of thousands of compounds (Gorgulla et al., 2022). The two primary categories of virtual screening are ligand-based and structure-based virtual screening (Figure 2) (Singh et al., 2024; Parvatikar et al., 2023; Lin et al., 2022). By employing a hierarchical approach that combines ligand-based and structure-based virtual screening, researchers can sequentially apply filters to reduce the size of screening libraries to a manageable scale for experimental validation, yielding successful outcomes in numerous drug screening initiatives (Kumar and Zhang, 2015). Wei et al. used virtual screening in their investigation of transcriptome analysis to elucidate the regulatory mechanisms underlying the biosynthesis of warm tulipane terpenoids induced by methyl jasmonate. This approach, which involved the prediction of protein analyses and Hidden Markov Model mapping of terpenes, provided a faster and more efficient way of studying regulatory mechanisms than conventional methodologies (Wei et al., 2022).

Recently, particularly following the emergence of the novel coronavirus in 2019, vaccine and drug development timelines have been prolonged. In this unique context, the integration of AI in drug screening has garnered significant attention from researchers and scholars in the field of pharmacology. To address the challenges associated with drug screening, selecting appropriate algorithmic models to enhance the screening capabilities of AI has become a focal point of interest within the research community.

3.1 AI in ligand-based virtual screening

The ligand-based approach involves analyzing established active molecules and the prediction of their pharmacological properties by evaluating the resemblance of the compound under examination to the recognized active compounds (Liu R. et al., 2021). Estimating the binding affinity between a ligand and its target, known as drug-target binding affinity, is a crucial stage in ligand-based virtual screening. The objective of ligand-based virtual screening is to identify molecules with unique fundamental structures with similar or superior biologically active ligands compared to known biochemical activity when a known active ligand binds to the drug target, a phenomenon referred to as "scaffold hopping" (Böhm et al., 2004). Ligand-based AI techniques can predict not only the biological activity of compounds but also their physicochemical and pharmacokinetic characteristics.

3.1.1 Predicting the biological activity of compounds

Determination of drug-target binding affinity is crucial for developing small-molecule drugs, as it indicates the strength of the interaction between a drug and its target (Sadybekov et al., 2022). Lack of affinity or interaction with non-target proteins can lead to ineffective therapeutic outcomes or potential toxicity. AI can help predict drug-target interactions by comparing drug and target characteristics to predict interactions, especially between similar drugs and targets (Öztürk et al., 2018). Currently, advanced models, including sequence-based, graph-based, and multimodal data-based

models, are used; however, they face challenges in mining edge information, acquiring pharmacophore knowledge, integrating multimodal data, and simulating interactions (Zhang et al., 2023). To overcome these limitations, Li Zhang et al. introduced a method called graphical features and pharmacophore-enhanced cross-attention networks for drug-target binding affinity prediction. This method employs graph neural network (GNN) modules, linear projection units, and self-attention layers to extract features of drugs and proteins. Additionally, intramolecular and intermolecular cross-attention mechanisms were designed to merge and interact with the drug and protein features. Linear projection units were used to obtain the final drug and protein features, and the binding affinity of the drug to the target was predicted using a multilayer perceptron (Zhang et al., 2023).

3.1.2 Predicting the physicochemical properties of compounds

Physicochemical characteristics indirectly influence the pharmacokinetic attributes and target receptor families of pharmaceuticals, making them crucial in the discovery and development of novel drugs (Zang et al., 2017). Currently, most techniques used to evaluate the physicochemical properties of drugs are based on traditional trial-and-error approaches. However, more sophisticated methods have been introduced as a result of recent advancements, such as six predictive models (SVMs, ANNs, lncRNA-disease associations, probabilistic neural network algorithms, kNN algorithms, and partial least squares) developed by researchers like Kumar (Kumar and Kim, 2021). These models were based on cumulative parameters like molecular refractive index, molecular volume, molecular surface area, molecular weight, log P, log S, and total polar surface area and have been used to predict the intestinal absorption of 497 compounds.

3.1.3 Predicting the absorption, distribution, metabolism, excretion, and toxicity tolerance (ADMET) properties of compounds

A successful drug should not only exhibit strong biological activity and favorable physicochemical characteristics but also possess excellent ADMET properties and undergo efficient pharmacokinetic processes. The inadequate pharmacokinetic attributes and potential toxicity of candidate compounds are significant contributors to the failure of drug development endeavors. The efficacy of targeted cancer therapies, a burgeoning antitumor treatment modality, necessitates drug delivery systems with minimal immunogenicity and toxicity levels (Luo et al., 2021). AI can be used to predict drug toxicity by analyzing the chemical structures and properties of compounds. ML algorithms, trained on toxicology databases, can predict detrimental effects and identify hazardous structural attributes. This predictive capacity aids researchers in prioritizing safer chemicals and mitigating adverse effects during clinical trials. For instance, Xiong et al. (2021) introduced the ADMETlab model, founded on *in silico* ADMET and is based on version 2.0 constructed using the Python Web framework Django. This model, hosted on the AliCloud Ubuntu Linux system, offers an expanded range of ADMET endpoints than its predecessor, including 17 physical chemistry, 13 medicinal chemistry, 23 ADME characterization, 27 toxicity endpoints, and 8 toxicogenic rules.

3.2 Utilization of AI in structure-based virtual screening

The application of AI in structure-based virtual screening, also known as target-based virtual screening, has become prevalent in the prediction of target-ligand interactions (Parvatikar et al., 2023). Structure-based virtual screening involves using molecular docking based on the three-dimensional (3D) structure of proteins to analyze the characteristics of target protein-binding sites and their interactions with small-molecule drugs. The binding affinity of the proteins to the drugs was evaluated using affinity scoring functions. Drugs exhibiting high predictive scores were selected from a vast pool of compound molecules for subsequent bioactivity testing. Molecular docking is the primary method used in this process (Singh et al., 2024). It is common to combine molecular docking with AI algorithms to validate docking outcomes and further refine compound screening (Lin et al., 2022). In contrast to ligand-based virtual screening, structure-based virtual screening can identify ligands with novel scaffolds or chemical functional groups (Dilip et al., 2016). For instance, fragments derived from natural products can be valuable lead compounds for the design and discovery of small-molecule drugs (Bai et al., 2021). The modification or simplification of natural product scaffolds is a key strategy in natural product drug discovery, and structure-based virtual screening can provide novel insights into this domain (Hui et al., 2024).

3.2.1 Molecular docking

Molecular docking is a computational technique used to predict the preferred orientation and structure of a ligand molecule when it interacts with another molecule, typically a larger receptor (Cerqueira et al., 2015). For instance, DeepDock, an AI-driven molecular docking approach, leverages deep learning algorithms to forecast binding modes and affinities in molecular docking scenarios. Unimol utilizes graph neural network technology to predict docking outcomes with enhanced efficiency. In addition, DiffDock amalgamates deep learning with graphical convolutional networks, thereby augmenting the accuracy and velocity of docking predictions (McNutt et al., 2021; He et al., 2024; B Fortela et al., 2024). This method is commonly used to predict target-ligand interactions (Pinzi and Rastelli, 2019). For instance, Yan et al. conducted molecular docking using Discovery Studio to evaluate the binding affinity of Mauritania with its potential targets, identifying PIK3CA as a promising target (Yan et al., 2022). In another study, Discovery Studio was used for molecular docking to identify the potential binding site of SIRT3 with craniospermone, revealing novel roles and mechanisms of craniospermone in anti-glomerular basement membrane antibody disease (GBM) and suggesting a new therapeutic approach for GBM treatment (Sun et al., 2022). Additionally, Gao Y et al. conducted molecular docking studies using the crystal structure of hHDAC1 to investigate the binding modes of compounds 27f and 39f to HDAC1 and HDAC6, demonstrating that the synthesized HDAC inhibitors 27f and 39f exhibited favorable binding modes and elucidated their potent inhibitory effects against HDAC1 and HDAC6 (Gao et al., 2023).

3.2.2 Predicting target structure

The prediction of the target structure plays a pivotal role in drug screening. Using AI, target protein structures can be predicted and evaluated to provide crucial insights for drug screening, thereby improving the efficiency of research and development endeavors. In a study of N⁶-methyladenosine (m⁶A), Liu et al. (Sui et al., 2020; Liu S. et al., 2023)established a novel database named "M6AREG" to facilitate the screening of drug-target interactions. They used crosslinking techniques and other methodologies to improve data collection, aiming to advance the future development of m⁶A research. Additionally, Can et al. identified potential peptide biomarkers within the Dendrobium genus using an AI-driven multivariate statistical analysis approach (Fang et al., 2020). Google's DeepMind introduced AlphaFold, a tool that leverages AI technology to train on Protein Data Bank structural data to predict the 3D structure of amino acid sequences (Powles and Hodson, 2017). The updated version of AlphaFold, AlphaFold2, integrates a novel graph neural network to improve the accuracy of target protein structure prediction (Lv et al., 2023). Although this method offers advantages in predicting target structures, there are areas of uncertainty (Meller et al., 2023). Therefore, new models, like the AlphaFold-multimer model, have been developed and used to significantly improve the accuracy of predicting multichain protein structures (Lv et al., 2023). Furthermore, the development of the MULTICOM four-level structure prediction system improved the AlphaFold-multimer model's ability to predict structurally intricate proteins (Liu J. et al., 2023). The latest iteration AlphaFold 3, which has a notable enhancement in DL algorithms and model architecture over its predecessors, allows for more effective management of intricate protein structures and extensive datasets. It is anticipated to achieve greater precision in the 3D structure prediction of proteins, particularly those that are complex and varied. The refined algorithms and computational infrastructure of AlphaFold 3 are designed to substantially decrease the prediction time. Additionally, the platform now encompasses advanced functionalities such as the prediction of protein-protein interactions and the binding of proteins to ligands (Jumper et al., 2021). However, protein folding remains a challenge for this approach (David et al., 2022). DN-fold, a DL technique for predicting protein-folded structures, has been proposed to address the prediction challenges encountered by AlphaFold (Jo et al., 2015). These examples underscore the notable achievements of AI-based target structure prediction, which has the potential to significantly improve the efficacy of drug discovery and development processes.

3.2.3 Predicting target-ligand interactions

The interactions between targets and ligands are intricate, and investigating their operational patterns can establish a theoretical foundation for drug screening. Using AI technology to analyze, assimilate, and prognosticate extensive datasets concerning known drug-receptor interactions can facilitate a comprehensive understanding of the mechanism of action and influential variables, thereby ensuring a logical and efficient drug screening process. Scoring functions are essential for predicting the binding affinities of drugs and targets (Selvaraj et al., 2022). The DL-based scoring function hinges on extracting features (such as distance and charge) from a representation of receptor-ligand interactions and subsequently predicting the binding affinity between the two entities (Kumar and Kim, 2021). Several scholars have combined RF and AutoDock scoring functions with the Glide XP score to

achieve superior scoring outcomes (Brown et al., 2021). Jiménez et al. (2018) developed a 3D graph-based CNN model to predict the interaction between ligand and receptor proteins, with their projected binding affinities closely aligning with the experimental data. Wang et al. developed a multiscale convolutional model capable of capturing the receptor and ligand characteristics to predict their interactions (Wang et al., 2021). Furthermore, Jin et al. (2021) introduced an EmbedDTI model, which improved the depiction of ligands and receptors while bolstering drug-target interaction prediction.

In the field of protein-protein interactions, numerous potential drug targets exist, and the process of discovering and identifying these targets is crucial for drug screening. For instance, in cancer therapeutic research, the use of SwissTargetPrediction for Muritan target prediction led to the identification of calmodulin as a promising target for lung cancer treatment (Chen et al., 2020). Furthermore, advancements in the high-resolution crystal structures of KDM1A inhibitor complexes have provided researchers with valuable insights into protein-ligand interactions, facilitating the screening of inhibitors with enhanced selectivity (He et al., 2022). Additionally, by investigating natural products, including flavonoids, alkaloids, and terpenoids, and their interactions with target proteins, scientists have gained insights into their ability to selectively target and inhibit inflammatory mediators, leading to the development of effective anti-inflammatory strategies (Li et al., 2023).

4 AI-powered drug design

The use of ML and DL algorithms in the field of drug development can facilitate the understanding of the intricate interplay between chemical and biological data (Mervin et al., 2021). Recently, computer-aided drug design (CADD) has gained significant traction in the pursuit of novel pharmaceuticals (Singh et al., 2024). Leveraging the computational capabilities of these algorithms has broadened their accessibility and reduced their costs. For example, Jiang XY et al. used CADD to investigate the structural modifications, monosubstituted derivatives, disubstituted derivatives, and disubstituted polymers of β-elemene, leading to improved hydrophilicity and antitumor activity. In addition to predicting the target structures and receptor-ligand interactions, CADD facilitates the *de novo* design of novel active compounds and the automated synthesis of drugs, thereby addressing challenges in organic synthesis (Jiang et al., 2024).

4.1 Redesigning active molecules

Creating novel active molecules from scratch represents a significant advancement in drug discovery and development. Computer-aided *ab initio* design to develop new active molecules has emerged as a valuable tool for drug design innovation (Reker et al., 2014). This approach involves using computational simulations and ML techniques to design bioactive molecular structures from first principles, aiming to address the scarcity of new chemical entities in drug discovery and cater to the demands of effective drug therapy (Reker et al., 2014).

Currently, there have been significant advancements in molecular design methodologies, including the use of autoencoders, GANs, and RNNs. Among these techniques, the variant autoencoder method stands out, comprising an encoder network and a decoder network (Gómez-Bombarelli et al., 2018). This method facilitates the transformation of chemical structures represented by the simplified molecular input line entry system (SMILES) notation into continuous real-valued vectors through the decoder. These continuous vectors serve as a potential space, with the decoder converting them back into chemical structures (Hessler and Baringhaus, 2018). Researchers have successfully used this method to train a model based on the quality estimation of drug-likeness ratings and synthetic accessibility score, generating more targeted molecules (Hessler and Baringhaus, 2018).

Furthermore, Kadurin et al. (2017) used GANs to propose novel molecules with potential anticancer properties. Additionally, RNNs have demonstrated promise in the *ab initio* design of drugs. RNNs are trained to encode chemical structures by generating numerous SMILES strings through training on a variety of compounds from different compound libraries. This approach has demonstrated the ability to generate novel peptide structures, demonstrating the potential of RNNs in drug design innovation (Hessler and Baringhaus, 2018).

4.2 Addressing the difficulties in organic synthesis and achieving the automation of drug synthesis

Addressing of challenges encountered in organic synthesis is a significant obstacle in the field of drug development. AI technology is crucial in predicting the reaction processes and optimizing synthetic pathways to facilitate rapid organic synthesis and drug design. The ML-based methods for forward synthesis prediction can accurately determine the sequence of synthetic routes and predict reactions and products (Hessler and Baringhaus, 2018). For example, PathPred (Table 1) is a robust web server capable of predicting multi-step synthetic pathways for a given compound, thereby enhancing the efficiency and speed of organic synthesis (Molnar et al., 2022). Liu et al. developed a model for retrosynthesis prediction that uses ML techniques and extensive datasets to accurately predict chemical reactions, thereby addressing the challenges of organic synthesis (Liu et al., 2017). Furthermore, AI enables the automation of compound synthesis (Mouchlis et al., 2021), as demonstrated by the development of Chemputer (Tripathi M. K. et al., 2021) a computer-aided software designed for automated compound synthesis.

5 AI's revolutionary role in drug clinical trials

Clinical trials are crucial for evaluating the safety, efficacy, and reliability of drug development processes. These trials follow a standardized, sequential methodology where scientists assess the safety, efficacy, and clinical relevance of promising new drugs (Bhavya et al., 2023; Vasan et al., 2023). However, clinical trials are labor-intensive and involve patient recruitment, enrollment, continuous monitoring, medical adherence, and data retention.

TABLE 1 AI based drug discovery and development tools and software.

Number	Drug discovery software based on AI	Auxiliary direction	Advantage
1	DeepDrug (Yin et al., 2023)	Detecting the genome of pathogens	Can access very large datasets and quickly identify new compounds
2	Reinvent (Blaschke et al., 2020)	Mainly used for molecular generation	Can conduct AI molecular design
3	AlphaFold (Nussinov et al., 2022)	Predicting the structure of proteins	Help understand the interaction between drugs and targets
4	SwissADME (Daina et al., 2017)	Predicting the oral bioavailability of compounds, etc	Quickly screen potential drug candidates in the early stages of drug development
5	ADMETSAR (Yang et al., 2018)	Predicting the toxicity of drug molecules	Plays an important role in drug development and safety evaluation
6	AutoDock Vina (Santos-Martins et al., 2019)	Determine binding conformation and binding affinity	Having high accuracy and relatively high docking speed
7	ADMETsar (Cheng et al., 2019)	Predicting the properties of compounds	Efficient, fast, and can reduce repeated experiments
8	DeepPurpose (Huang et al., 2021)	Drug activity prediction	Easy to operate and flexible to use
19	MATLAB (Molnar et al., 2022)	Assist in exploring the potential characteristics of drugs	The algorithm is efficient and highly targeted
10	PathPred (Ma et al., 2020)	Predicting the synthetic route of compounds	Can conveniently and intuitively demonstrate the synthetic route of compounds
11	Deep 6 AI (Lei et al., 2024; Liu et al., 2022)	Assist in quickly identifying potential trial participants	Improve patient recruitment efficiency

Personalized AI solutions can streamline and expedite these experiments by managing trial data, integrating patient histories, and focusing on patient-centered AI approaches (Chopra et al., 2023). The advent of AI has revolutionized the data collection and monitoring aspects of clinical trials, leading to reduced costs, increased efficiency, and improved drug development research. The ways in which AI can contribute to pharmacy clinical trials, including its applications in participant recruitment, data collection and analysis, predictive analytics for trial design, and patient monitoring and safety, were summarized in this study.

5.1 Improving participant recruitment efficiency

In phase I clinical trials, approximately 80% of trials experience delays in patient enrollment (Chopra et al., 2023). The recruitment of suitable trial participants is a time-consuming and costly aspect of clinical trials. Traditional methods of participant recruitment in clinical trials involve professionals manually screening extensive medical records, posing challenges in terms of both quantity and quality (Lovato et al., 1997; Britton et al., 1999). Conversely, AI technology can quickly identify potential participants meeting specific trial criteria by analyzing electronic health records, social media platforms, and other online data sources. For instance, Deep 6 AI (Table 1) has developed a technology capable of sifting through millions of patient records to quickly identify suitable clinical trial candidates. Similarly, Mendel. AI uses an AI system for precise matching of medical records with clinical trial data, ensuring prompt notification to patients to promote diversity and representativeness in study samples (Lei et al., 2024; Liu et al., 2022).

5.2 Improving data collection and analysis

Conventional methods for collecting and evaluating data from drug clinical trials exhibit various limitations, including reduced data efficiency, constrained utilization, susceptibility to errors, limited scalability, and inadequate monitoring (Jüni et al., 1999). Conversely, AI provides a solution to these challenges. AI plays a crucial role in improving the quality and expediency of data collection and analysis in drug clinical trials by effectively processing and conducting in-depth analyses of extensive and intricate datasets. ML techniques excel in managing and analyzing large datasets from clinical trials, enabling the identification of overlooked issues and concealed risks. Furthermore, AI can monitor real-time data from wearable devices to monitor participants' health metrics, thereby providing researchers with accurate and timely information (Lei et al., 2024; Tripathi N. et al., 2021; Kolluri et al., 2022)

5.3 Predictive analytics applied to trial design

AI-powered predictive analytics are instrumental in the trial design phase as they aid in predicting the potential outcomes of a clinical trial based on the analysis of historical trial data and other relevant information (Fleming and DeMets, 1996). By leveraging AI, researchers can predict the efficacy of drug treatments before conducting trials, enabling informed decisions regarding trial design, such as determining the optimal drug dosage and the early detection of potential side effects (Aliper et al., 2023).

5.4 Improved patient monitoring and safety measures

Furthermore, AI-based monitoring tools are crucial in improving patient safety by automatically collecting data to promptly identify and address safety concerns. Moreover, AI technology can assist in monitoring patient adherence to trial protocols, thereby ensuring the credibility and accuracy of trial outcomes (Moazemi et al., 2023).

6 Challenges of AI in drug development

AI has demonstrated improved outcomes across various stages of drug development. However, significant challenges remain in this domain. First, the intricate nature of drug actions within an organism presents a complex hurdle (Yu et al., 2023). The failure of certain drugs to progress to clinical trials can be attributed to the fact that AI-driven drug development processes are conducted in a controlled environment that lacks the complexity of real-world conditions (Yu et al., 2023). Second, the efficacy of AI models in drug screening is influenced by the quality and diversity of research data. Moreover, AI relies significantly on data and specialized personnel. Additionally, concerns arise when technologies like facial recognition software or other ML tools are used to monitor trial participants, potentially infringing upon their privacy (Almeida et al., 2022). Furthermore, AI is susceptible to issues like programming errors that can lead to missed opportunities in drug studies. These challenges are not easily surmountable in the short term. The integration of AI in drug research not only tests the limits of time and technology but also represents humanity's quest into uncharted territories, signifying a long and arduous journey ahead.

7 Prospect

Integration of AI is becoming more common in the field of drug development. The current era is characterized by a surge in AI technologies, exemplified by the introduction of the text-to-image model Dall-E in 2021 and its successor Dall-E2 in 2022 (Adams et al., 2023). These tools are considered promising for image generation, enhancement, and manipulation in future radiology AI research. The release of ChatGPT was succeeded by GPT-4, demonstrating its significant potential for application in pharmacology, spanning from research topic identification to clinical laboratory diagnosis (Dave et al., 2023). Recently, OpenAI formally introduced the text-to-video model Sora, which is currently in the feedback acquisition phase. Despite being in the early stages of development, Sora's performance and potential are comparable to that of GPT-4. It is anticipated that Sora will soon be used in drug discovery. The integration of AI technology into drug development and mining endeavors is poised to make a significant and groundbreaking impact on human health and wellbeing.

Author contributions

YW: Investigation, Writing–original draft. LM: Investigation, Writing–original draft. XL: Investigation, Writing–original draft. JY: Investigation, Writing–review and editing. XR: Investigation, Writing–review and editing. YH: Investigation, Writing–review and editing. JX: Investigation, Writing–review and editing. LT: Investigation, Writing–review and editing, Supervision. JW: Investigation, Writing–review and editing. LD: Supervision, Investigation, Writing–review and editing. GC: Investigation, Supervision, Writing–review and editing. SL: Writing–review and editing, Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Validation.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Adams, L. C., Busch, F., Truhn, D., Makowski, M. R., Aerts, H. J. W. L., and Bressem, K. K. (2023). What does DALL-E 2 know about radiology? *J. Med. Internet Res.* 25, e43110. doi:10.2196/43110

Aliper, A., Kudrin, R., Polykovskiy, D., Kamya, P., Tutubalina, E., Chen, S., et al. (2023). Prediction of clinical trials outcomes based on target choice and clinical trial design with multi-modal artificial intelligence. *Clin. Pharmacol. Ther.* 114 (5), 972–980. doi:10.1002/cpt.3008

Almeida, D., Denise, K., and Lomas, E. (2022). The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. *AI ethics* 2 (3), 377–387. doi:10.1007/s43681-021-00077-w

Bai, R., Yao, C., Zhong, Z., Ge, J., Bai, Z., Ye, X., et al. (2021). Discovery of natural anti-inflammatory alkaloids: potential leads for the drug discovery for the treatment of inflammation. *Eur. J. Med. Chem.* 213, 113165. doi:10.1016/j.ejmech.2021.113165

B Fortela, D. L., Mikolajczyk, A. P., Carnes, M. R., Sharp, W., Revellame, E., Hernandez, R., et al. (2024). Predicting molecular docking of per- and polyfluoroalkyl substances to blood protein using generative artificial intelligence algorithm DiffDock. *Biotechniques* 76 (1), 14–26. doi:10.2144/btn-2023-0070

Bhavya, K. S., and Sheth, D. K. (2023). "Chapter 66-Phases of clinical trails," in *Handbook for designing and conducting clinical and translational research, translational sports medicine*. Editors A. E. M. Eltorai, J. A. Bakal, S. F. DeFroda, and B. D. Owens (Academic Press), 331–333. doi:10.1016/B978-0-323-91259-4.00006-0

Blaschke, T., Josep, A., Chen, H., Margreitter, C., Tyrchan, C., Engkvist, O., et al. (2020). REINVENT 2.0: an AI tool for *de novo* drug design. *J. Chem. Inf. Model.* 60 (12), 5918–5922. doi:10.1021/acs.jcim.0c00915

Böhm, H. J., Flohr, A., and Stahl, M. (2004). Scaffold hopping. *Drug Discov. Today Technol.* 1 (3), 217–224. doi:10.1016/j.ddtec.2004.10.009

Britton, A., McKee, M., Black, N., McPherson, K., Sanderson, C., and Bain, C. (1999). Threats to applicability of randomised trials: exclusions and selective participation. *J. Health Serv. Res. Policy* 4 (2), 112–121. doi:10.1177/135581969900400210

Brown, B. P., Mendenhall, J., Geanes, A. R., and Meiler, J. (2021). General purpose structure-based drug discovery neural network score functions with human-interpretable pharmacophore maps. *J. Chem. Inf. Model* 61 (2), 603–620. doi:10.1021/acs.jcim.0c01001

Cerqueira, N. M., Gesto, D., Oliveira, E. F., Santos-Martins, D., Brás, N. F., Sousa, S. F., et al. (2015). Receptor-based virtual screening protocol for drug discovery. *Arch. Biochem. Biophys.* 582, 56–67. doi:10.1016/j.abb.2015.05.011

Chen, P., Wu, Q., Feng, J., Yan, L., Sun, Y., Liu, S., et al. (2020). Erianin, a novel dibenzyl compound in Dendrobium extract, inhibits lung cancer cell growth and migration via calcium/calmodulin-dependent ferroptosis. *Signal Transduct. Target Ther.* 5 (1), 51. doi:10.1038/s41392-020-0149-3

Chen, R., Wei, Q., Liu, Y., Wei, X., Chen, X., Yin, X., et al. (2021). Transcriptome sequencing and functional characterization of new sesquiterpene synthases from Curcuma wenyujin. *Arch. Biochem. Biophys.* 709, 108986. doi:10.1016/j.abb.2021.108986

Cheng, F., Li, W., Zhou, Y., Shen, J., Wu, Z., Liu, G., et al. (2019). admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties. *J. Chem. Inf. Model* 59 (11), 4959. doi:10.1021/acs.jcim.9b00969

Chopra, H., Annu, Shin, D. K., Munjal, K., Priyanka, Dhama, K., et al. (2023). Revolutionizing clinical trials: the role of AI in accelerating medical breakthroughs. *Int. J. Surg.* 109 (12), 4211–4220. doi:10.1097/JS9.0000000000000705

Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7, 42717. doi:10.1038/srep42717

Dave, T., Athaluri, S. A., and Singh, S. (2023). ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front. Artif. Intell.* 6, 1169595. doi:10.3389/frai.2023.1169595

David, A., Islam, S., Tankhilevich, E., and Sternberg, M. J. E. (2022). The AlphaFold database of protein structures: a biologist's guide. *J. Mol. Biol.* 434 (2), 167336. doi:10.1016/j.jmb.2021.167336

Dilip, A., Lešnik, S., Štular, T., Janežič, D., and Konc, J. (2016). Ligand-based virtual screening interface between PyMOL and LiSiCA. *J. Cheminform* 8 (1), 46. doi:10.1186/s13321-016-0157-z

Dueñas, M. E., Peltier-Heap, R. E., Leveridge, M., Annan, R. S., Büttner, F. H., and Trost, M. (2023). Advances in high-throughput mass spectrometry in drug discovery. *EMBO Mol. Med.* 15 (1), e14850. doi:10.15252/emmm.202114850

Fang, C., Xin, G. Z., Wang, S. L., Wei, M. M., Wu, P., Dong, X. M., et al. (2020). Discovery and validation of peptide biomarkers for discrimination of Dendrobium species by label-free proteomics and chemometrics. *J. Pharm. Biomed. Anal.* 182, 113118. doi:10.1016/j.jpba.2020.113118

Fleming, T. R., and DeMets, D. L. (1996). Surrogate end points in clinical trials: are we being misled? *Ann. Intern Med.* 125 (7), 605–613. doi:10.7326/0003-4819-125-7-199610010-00011

Gao, Y., Duan, J., Dang, X., Yuan, Y., Wang, Y., He, X., et al. (2023). Design, synthesis and biological evaluation of novel histone deacetylase (HDAC) inhibitors derived from β-elemene scaffold. *J. Enzyme Inhib. Med. Chem.* 38 (1), 2195991. doi:10.1080/14756366.2023.2195991

Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., et al. (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* 4 (2), 268–276. doi:10.1021/acscentsci.7b00572

Gorgulla, C., Jayaraj, A., Fackeldey, K., and Arthanari, H. (2022). Emerging frontiers in virtual drug discovery: from quantum mechanical methods to deep learning approaches. *Curr. Opin. Chem. Biol.* 69, 102156. doi:10.1016/j.cbpa.2022.102156

Guan, Q. L. Y., Liu, Y., Li, Y., Yang, H., Li, Y., and Li, X. (2023). Application and progress of artificial intelligence in anti-tumor drug research and development. *Chin. Mod. Appl. Pharm.* 23, 3318–3323. doi:10.13748/j.carolcarrollnkiissn1007-7693.20230857

Gulakala, R., Markert, B., and Stoffel, M. (2022). Generative adversarial network based data augmentation for CNN based detection of Covid-19. *Sci. Rep.* 12 (1), 19186. doi:10.1038/s41598-022-23692-x

Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., and Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol. Divers* 25 (3), 1315–1360. doi:10.1007/s11030-021-10217-3

He, X., Zhang, H., Zhang, Y., Ye, Y., Wang, S., Bai, R., et al. (2022). Drug discovery of histone lysine demethylases (KDMs) inhibitors (progress from 2018 to present). *Eur. J. Med. Chem.* 231, 114143. doi:10.1016/j.ejmech.2022.114143

He, X., Zhao, L., Tian, Y., Li, R., Chu, Q., Gu, Z., et al. (2024). Highly accurate carbohydrate-binding site prediction with DeepGlycanSite. *Nat. Commun* 15 (1), 5163. doi:10.1038/s41467-024-49516-2

Hessler, G., and Baringhaus, K. H. (2018). Artificial intelligence in drug design. *Molecules* 23 (10), 2520. doi:10.3390/molecules23102520

Huang, K., Fu, T., Glass, L. M., Zitnik, M., Xiao, C., and Sun, J. (2021). DeepPurpose: a deep learning library for drug-target interaction prediction. *Bioinformatics* 36 (22-23), 5545–5547. doi:10.1093/bioinformatics/btaa1005

Hui, Z., Wen, H., Zhu, J., Deng, H., Jiang, X., Ye, X. Y., et al. (2024). Discovery of plant-derived anti-tumor natural products: potential leads for anti-tumor drug discovery. *Bioorg Chem.* 142, 106957. doi:10.1016/j.bioorg.2023.106957

Islam, M. A., Hasan, M. M. Z., Hussein, M. A., Hossain, K. M., and Miah, M. S. (2024). A review of machine learning and deep learning algorithms for Parkinson's disease detection using handwriting and voice datasets. *Heliyon* 10 (3), e25469. doi:10.1016/j.heliyon.2024.e25469

Jiang, X. Y., Shi, L. P., Zhu, J. L., Bai, R. R., and Xie, T. (2024). Elemene antitumor drugs development based on "molecular compatibility theory" and clinical application: a retrospective and prospective outlook. *Chin. J. Integr. Med.* 30 (1), 62–74. doi:10.1007/s11655-023-3714-0

Jiménez, J., Škalič, M., Martínez-Rosell, G., and De Fabritiis, G. (2018). K_{DEEP}: protein-ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J. Chem. Inf. Model* 58 (2), 287–296. doi:10.1021/acs.jcim.7b00650

Jin, Y., Lu, J., Shi, R., and Yang, Y. (2021). EmbedDTI: enhancing the molecular representations via sequence embedding and graph convolutional network for the prediction of drug-target interaction. *Biomolecules* 11 (12), 1783. doi:10.3390/biom11121783

Jo, T., Hou, J., Eickholt, J., and Cheng, J. (2015). Improving protein fold recognition by deep learning networks. *Sci. Rep.* 5, 17573. doi:10.1038/srep17573

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596 (7873), 583–589. doi:10.1038/s41586-021-03819-2

Jüni, P., Witschi, A., Bloch, R., and Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 282 (11), 1054–1060. doi:10.1001/jama.282.11.1054

Kadurin, A., Aliper, A., Kazennov, A., Mamoshina, P., Vanhaelen, Q., Khrabrov, K., et al. (2017). The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* 8 (7), 10883–10890. doi:10.18632/oncotarget.14073

Kaul, V., Enslin, S., and Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointest. Endosc.* 92 (4), 807–812. doi:10.1016/j.gie.2020.06.040

Kolluri, S., Lin, J., Liu, R., Zhang, Y., and Zhang, W. (2022). Machine learning and artificial intelligence in pharmaceutical research and development: a review. *AAPS J.* 24 (1), 19. doi:10.1208/s12248-021-00644-3

Kumar, A., and Zhang, K. Y. (2015). Hierarchical virtual screening approaches in small molecule drug discovery. *Methods* 71, 26–37. doi:10.1016/j.ymeth.2014.07.007

Kumar, S., and Kim, M. H. (2021). SMPLIP-Score: predicting ligand binding affinity from simple and interpretable on-the-fly interaction fingerprint pattern descriptors. *J. Cheminform* 13 (1), 28. doi:10.1186/s13321-021-00507-1

Lac, L., Leung, C. K., and Hu, P. (2024). Computational frameworks integrating deep learning and statistical models in mining multimodal omics data. *J. Biomed. Inf.* 152, 104629. doi:10.1016/j.jbi.2024.104629

Lei, F., Du, L., Dong, M., and Liu, X. M. (2024). Transparent report of the early-stage clinical evaluation of clinical Decision support system based on artificial intelligence. *Chin. general Med.* 27 (10), 1267–1270. doi:10.12114/j.iSSN.1007-9572.2023.0668

Lei, Y., Li, S., Liu, Z., Wan, F., Tian, T., Li, S., et al. (2021). A deep-learning framework for multi-level peptide-protein interaction prediction. *Nat. Commun.* 12 (1), 5465. doi:10.1038/s41467-021-25772-4

Li, J., Zhao, R., Miao, P., Xu, F., Chen, J., Jiang, X., et al. (2023). Discovery of anti-inflammatory natural flavonoids: diverse scaffolds and promising leads for drug discovery. *Eur. J. Med. Chem.* 260, 115791. doi:10.1016/j.ejmech.2023.115791

Lim, S., Lee, K., and Kang, J. (2018). Drug drug interaction extraction from the literature using a recursive neural network. *PLoS One* 13 (1), e0190926. doi:10.1371/journal.pone.0190926

Lin, Y., Zhang, Y., Wang, D., Yang, B., and Shen, Y. Q. (2022). Computer especially AI-assisted drug virtual screening and design in traditional Chinese medicine. *Phytomedicine* 107, 154481. doi:10.1016/j.phymed.2022.154481

Liu, B., Ramsundar, B., Kawthekar, P., Shi, J., Gomes, J., Luu Nguyen, Q., et al. (2017). Retrosynthetic reaction prediction using neural sequence-to-sequence models. *ACS Cent. Sci.* 3 (10), 1103–1113. doi:10.1021/acscentsci.7b00303

Liu, C., Shin, J., Son, S., Choe, Y., Farokhzad, N., Tang, Z., et al. (2021a). Pnictogens in medicinal chemistry: evolution from erstwhile drugs to emerging layered photonic nanomedicine. *Chem. Soc. Rev.* 50 (4), 2260–2279. doi:10.1039/d0cs01175d

Liu, J., Guo, Z., Wu, T., Roy, R. S., Quadir, F., Chen, C., et al. (2023b). Enhancing alphafold-multimer-based protein complex structure prediction with MULTICOM in CASP15. *Commun. Biol.* 6 (1), 1140. doi:10.1038/s42003-023-05525-3

Liu, R., Song, J., Liu, A., and Du, G. (2021b). Application of artificial intelligence in drug screening based on ligand and receptor structure. *Acta Pharmacol. Sin.* 08, 2136–2145. doi:10.16438/j.0513-4870.2021-0052

Liu, S., Chen, L., Zhang, Y., Zhou, Y., He, Y., Chen, Z., et al. (2023a). M6AREG: m6A-centered regulation of disease development and drug response. *Nucleic Acids Res.* 51 (D1), D1333–D1344. doi:10.1093/nar/gkac801

Liu, X., Lu, X. R., Wu, Y., Yu, H. T., and Wang, X. M. (2022). Ethical analysis and countermeasures of artificial intelligence application in clinical trials. *Chin. J. Clin. Pharmacol. Ther.* 27 (3), 322–327. doi:10.12092/j.issn.1009-2501.2022.03.012

Lovato, L. C., Hill, K., Hertert, S., Hunninghake, D. B., and Probstfield, J. L. (1997). Recruitment for controlled clinical trials: literature summary and annotated bibliography. *Control Clin. Trials* 18 (4), 328–352. doi:10.1016/s0197-2456(96)00236-x

Luo, R., Liu, M., Tan, T., Yang, Q., Wang, Y., Men, L., et al. (2021). Emerging significance and therapeutic potential of extracellular vesicles. *Int. J. Biol. Sci.* 17 (10), 2476–2486. doi:10.7150/ijbs.59296

Lv, Q., Zhou, F., Liu, X., and Zhi, L. (2023). Artificial intelligence in small molecule drug discovery from 2018 to 2023: does it really work? *Bioorg Chem.* 141, 106894. doi:10.1016/j.bioorg.2023.106894

Ma, X., Meng, Y., Wang, P., Tang, Z., Wang, H., and Xie, T. (2020). Bioinformatics-assisted, integrated omics studies on medicinal plants. *Brief. Bioinform* 21 (6), 1857–1874. doi:10.1093/bib/bbz132

Mak, K. K., and Pichika, M. R. (2019). Artificial intelligence in drug development: present status and future prospects. *Drug Discov. Today.* 24 (3), 773–780. doi:10.1016/j.drudis.2018.11.014

Mao, Y., and Liu, P. (2021). Application and Innovation of artificial intelligence in drug research and development. *Chin. J. new drugs Clin.* (6), 430–435. doi:10.14109/j.carollcarrollnkixyylc.2021.06.07

McNutt, A. T., Francoeur, P., Aggarwal, R., Masuda, T., Meli, R., Ragoza, M., et al. (2021). GNINA 1.0: molecular docking with deep learning. *J. Cheminform* 13 (1), 43. doi:10.1186/s13321-021-00522-2

Meller, A., Bhakat, S., Solieva, S., and Bowman, G. R. (2023). Accelerating cryptic pocket discovery using AlphaFold. *J. Chem. Theory Comput.* 19 (14), 4355–4363. doi:10.1021/acs.jctc.2c01189

Mervin, L. H., Johansson, S., Semenova, E., Giblin, K. A., and Engkvist, O. (2021). Uncertainty quantification in drug design. *Drug Discov. Today* 26 (2), 474–489. doi:10.1016/j.drudis.2020.11.027

Mintz, Y., and Brodie, R. (2019). Introduction to artificial intelligence in medicine. *Minim. Invasive Ther. Allied Technol.* 28 (2), 73–81. doi:10.1080/13645706.2019.1575882

Moazemi, S., Vahdati, S., Li, J., Kalkhoff, S., Castano, L. J. V., Dewitz, B., et al. (2023). Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: a systematic review. *Front. Med. (Lausanne)* 10, 1109411. doi:10.3389/fmed.2023.1109411

Molnar, D., Röhm, M., Wutz, J., Rivec, I., Michel, A., Klotz, G., et al. (2022). A novel MATLAB®-Algorithm-Based video analysis to quantitatively determine solution creeping in intact pharmaceutical glass vials. *Eur. J. Pharm. Biopharm.* 178, 117–130. doi:10.1016/j.ejpb.2022.08.003

Mouchlis, V. D., Afantitis, A., Serra, A., Fratello, M., Papadiamantis, A. G., Aidinis, V., et al. (2021). Advances in *de novo* drug design: from conventional to machine learning methods. *Int. J. Mol. Sci.* 22 (4), 1676. doi:10.3390/ijms22041676

Nussinov, R., Zhang, M., Liu, Y., and Jang, H. (2022). AlphaFold, artificial intelligence (AI), and allostery. *J. Phys. Chem. B* 126 (34), 6372–6383. doi:10.1021/acs.jpcb.2c04346

Öztürk, H., Özgür, A., and Ozkirimli, E. (2018). DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics* 34 (17), i821–i829. doi:10.1093/bioinformatics/bty593

Pandiyan, S., and Wang, L. (2022). A comprehensive review on recent approaches for cancer drug discovery associated with artificial intelligence. *Comput. Biol. Med.* 150, 106140. doi:10.1016/j.compbiomed.2022.106140

Parvatikar, P. P., Patil, S., Khaparkhuntikar, K., Patil, S., Singh, P. K., Sahana, R., et al. (2023). Artificial intelligence: machine learning approach for screening large database and drug discovery. *Antivir. Res.* 220, 105740. doi:10.1016/j.antiviral.2023.105740

Pinzi, L., and Rastelli, G. (2019). Molecular docking: shifting paradigms in drug discovery. *Int. J. Mol. Sci.* 20 (18), 4331. doi:10.3390/ijms20184331

Powles, J., and Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health Technol. Berl.* 7 (4), 351–367. doi:10.1007/s12553-017-0179-1

Qi, X., Jiang, S. W., Yuan, Y. H., Xu, L., Hui, Z., Ye, X. Y., et al. (2022). Research progress of adaptor associated protein kinase 1(AAK1) inhibitors against virus. *Chin. J. Med.* 57(7), 1991–2002. doi:10.16438/j.0513-4870.2022-0155

Reker, D., Rodrigues, T., Schneider, P., and Schneider, G. (2014). Identifying the macromolecular targets of *de novo*-designed chemical entities through self-organizing map consensus. *Proc. Natl. Acad. Sci. U. S. A.* 111 (11), 4067–4072. doi:10.1073/pnas.1320001111

Rigatti, S. J. (2017). Random forest. *J. Insur. Med.* 47 (1), 31–39. doi:10.17849/insm-47-01-31-39.1

Ryu, J. Y., Lee, J. H., Lee, B. H., Song, J. S., Ahn, S., and Oh, K. S. (2022). PredMS: a random forest model for predicting metabolic stability of drug candidates in human liver microsomes. *Bioinformatics* 38 (2), 364–368. doi:10.1093/bioinformatics/btab547

Sadybekov, A. A., Sadybekov, A. V., Liu, Y., Iliopoulos-Tsoutsouvas, C., Huang, X. P., Pickett, J., et al. (2022). Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* 601 (7893), 452–459. doi:10.1038/s41586-021-04220-9

Santos-Martins, D., Eberhardt, J., Bianco, G., Solis-Vasquez, L., Ambrosio, F. A., Koch, A., et al. (2019). D3R Grand Challenge 4: prospective pose prediction of BACE1 ligands with AutoDock-GPU. *J. Comput. Aided Mol. Des.* 33 (12), 1071–1081. doi:10.1007/s10822-019-00241-9

Selvaraj, C., Chandra, I., and Singh, S. K. (2022). Artificial intelligence and machine learning approaches for drug design: challenges and opportunities for the pharmaceutical industries. *Mol. Divers* 26 (3), 1893–1913. doi:10.1007/s11030-021-10326-z

Shi, H., Tian, S., Li, Y., Li, D., Yu, H., Zhen, X., et al. (2015). Absorption, distribution, metabolism, excretion, and toxicity evaluation in drug discovery. 14. Prediction of human pregnane X receptor activators by using naive bayesian classification technique. *Chem. Res. Toxicol.* 28 (1), 116–125. doi:10.1021/tx500389q

Shiammala, P. N., Duraimutharasan, N. K. B., Vaseeharan, B., Alothaim, A. S., Al-Malki, E. S., Snekaa, B., et al. (2023). Exploring the artificial intelligence and machine learning models in the context of drug design difficulties and future potential for the pharmaceutical sectors. *Methods* 219, 82–94. doi:10.1016/j.ymeth.2023.09.010

Singh, S., Gupta, H., Sharma, P., and Sahi, S. (2024). Advances in Artificial Intelligence (AI)-assisted approaches in drug screening. *Artif. Intell. Chem.* 18 (3), 100039–100313. doi:10.1016/j.aichem.2023.100039

Smith, L. A., Oakden-Rayner, L., Bird, A., Zeng, M., To, M. S., Mukherjee, S., et al. (2023). Machine learning and deep learning predictive models for long-term prognosis in patients with chronic obstructive pulmonary disease: a systematic review and meta-analysis. *Lancet Digit. Health* 5 (12), e872–e881. doi:10.1016/S2589-7500(23)00177-2

Sui, X., Zhang, M., Han, X., Zhang, R., Chen, L., Liu, Y., et al. (2020). Combination of traditional Chinese medicine and epidermal growth factor receptor tyrosine kinase inhibitors in the treatment of non-small cell lung cancer: a systematic review and meta-analysis. *Med. Baltim.* 99 (32), e20683. doi:10.1097/MD.0000000000020683

Sun, S., Shi, J., Wang, X., Huang, C., Huang, Y., Xu, J., et al. (2022). Atractylon inhibits the tumorigenesis of glioblastoma through SIRT3 signaling. *Am. J. Cancer Res.* 12 (5), 2310–2322.

Tang, Z., Guo, J., Tang, J., and Ying, X. (2022). Analysis of current status of marketed innovative drugs in China from 2017 to 2020. *World Clin. Med.* (3), 294–299. doi:10.13683/j.ph.2022.03.015

Tripathi, M. K., Nath, A., Singh, T. P., Ethayathulla, A. S., and Kaur, P. (2021a). Evolving scenario of big data and Artificial Intelligence (AI) in drug discovery. *Mol. Divers* 25 (3), 1439–1460. doi:10.1007/s11030-021-10256-w

Tripathi, N., Goshisht, M. K., Sahu, S. K., and Arora, C. (2021b). Applications of artificial intelligence to drug design and discovery in the big data era: a comprehensive review. *Mol. Divers* 25 (3), 1643–1664. doi:10.1007/s11030-021-10237-z

Tuan, X. T., Ngan, D. K., and Huang, R. (2023) "Chapter 18-Application of QSAR models based on machine learning methods in chemical risk assessment and drug discovery," in *Huixiao hong, QSAR in safety evaluation and risk assessment*. Academic Press, 245–258. doi:10.1016/B978-0-443-15339-6.00006-0

Vasan, K., Gysi, D. M., and Barabási, A. L. (2023). The clinical trials puzzle: how network effects limit drug discovery. *iScience* 26 (12), 108361. doi:10.1016/j.isci.2023.108361

Wang, F., Feng, X., Kong, R., and Chang, S. (2023). Generating new protein sequences by using dense network and attention mechanism. *Math. Biosci. Eng.* 20 (2), 4178–4197. doi:10.3934/mbe.2023195

Wang, S., Jiang, M., Zhang, S., Wang, X., Yuan, Q., Wei, Z., et al. (2021). MCN-CPI: multiscale convolutional network for compound-protein interaction prediction. *Biomolecules* 11 (8), 1119. doi:10.3390/biom11081119

Wang, Y., Yu, X., Gu, Y., Li, W., Zhu, K., Chen, L., et al. (2024). XGraphCDS: an explainable deep learning model for predicting drug sensitivity from gene pathways and chemical structures. *Comput. Biol. Med.* 168, 107746. doi:10.1016/j.compbiomed.2023.107746

Wei, Q., Lan, K., Liu, Y., Chen, R., Hu, T., Zhao, S., et al. (2022). Transcriptome analysis reveals regulation mechanism of methyl jasmonate-induced terpenes biosynthesis in Curcuma wenyujin. *PLoS One* 17 (6), e0270309. doi:10.1371/journal.pone.0270309

Wu, H., Liu, J., Zhang, R., Lu, R., Cui, G., Cui, Z., et al. (2024). A review of deep learning methods for ligand based drug virtual screening. *Fundam. Res.* 4, 715–737. doi:10.1016/j.fmre.2024.02.011

Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., et al. (2021). ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 49 (W1), W5–W14. doi:10.1093/nar/gkab255

Yan, L., Zhang, Z., Liu, Y., Ren, S., Zhu, Z., Wei, L., et al. (2022). Anticancer activity of erianin: cancer-specific target prediction based on network pharmacology. *Front. Mol. Biosci.* 9, 862932. doi:10.3389/fmolb.2022.862932

Yang, H., Sun, L., Wang, Z., Li, W., Liu, G., and Tang, Y. (2018). ADMETopt: a web server for ADMET optimization in drug design via scaffold hopping. *J. Chem. Inf. Model* 58 (10), 2051–2056. doi:10.1021/acs.jcim.8b00532

Yang, M., Yang, B., Duan, G., and Wang, J. (2023). ITRPCA: a new model for computational drug repositioning based on improved tensor robust principal component analysis. *Front. Genet.* 14, 1271311. doi:10.3389/fgene.2023.1271311

Yin, Q. J., Fan, R., Cao, X. S., Liu, Q., Jiang, R., and Zeng, W. W. (2023). *DeepDrug: a general graph-based deep learning framework for drug-drug interactions and drug-target interactions prediction. Quant. Biol.*, 11 (3), 260–274. doi:10.15302/J-QB-022-0320

You, Y., Luo, L., You, Y., Lin, Y., Hu, H., Chen, Y., et al. (2020). Shengmai Yin formula modulates the gut microbiota of spleen-deficiency rats. *Chin. Med.* 15, 114. doi:10.1186/s13020-020-00394-y

Yu, Z., Zhang, L., Zhang, M., Dai, Z., Peng, C., and Zheng, S. (2023). Ai-based drug Discovery: current progress and future challenges. *J. China Pharm. Univ.* (03), 282–293. doi:10.11665/j.issn.1000-5048.2023041003

Yuan, Y., Duan, J., Huizi, , et al. (2022). Research progress of targeted ATR kinase inhibitors in cancer treatment. *Acta Pharmacol. Sin.* 57 (3), 593–604. doi:10.16438/J.0513-4870.2021-1522

Yuan, Y. H., Mao, N. D., Duan, J. L., Zhang, H., Garrido, C., Lirussi, F., et al. (2023). Recent progress in discovery of novel AAK1 inhibitors: from pain therapy to potential anti-viral agents. *J. Enzyme Inhib. Med. Chem.* 38 (1), 2279906. doi:10.1080/14756366.2023.2279906

Zang, Q., Mansouri, K., Williams, A. J., Judson, R. S., Allen, D. G., Casey, W. M., et al. (2017). *In silico* prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. *J. Chem. Inf. Model* 57 (1), 36–49. doi:10.1021/acs.jcim.6b00625

Zhang, J. F., Jiang, L. D., and Zhang, Y. L. (2014). Application of support vector machine in screening neurotoxic compounds from traditional Chinese medicine. 39(17), 3330–3334. doi:10.4268/cjcmm20141724

Zhang, L., Wang, C. C., Zhang, Y., and Chen, X. (2023). GPCNDTA: prediction of drug-target binding affinity through cross-attention networks augmented with graph features and pharmacophores. *Comput. Biol. Med.* 166, 107512. doi:10.1016/j.compbiomed.2023.107512

Zheng, M., and Jiang, H. (2019). High-value data mining and artificial intelligence technology accelerate innovative drug development [J]. *Adv. Pharm. Sci.*, 45(07), 481–483.