Taylor & Francis
Taylor & Francis Group

Expert Opinion on Drug Discovery

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/iedc20

Artificial intelligence for small molecule anticancer drug discovery

Lihui Duo, Yu Liu, Jianfeng Ren, Bencan Tang & Jonathan D. Hirst

To cite this article: Lihui Duo, Yu Liu, Jianfeng Ren, Bencan Tang & Jonathan D. Hirst (18 Jun 2024): Artificial intelligence for small molecule anticancer drug discovery, Expert Opinion on Drug Discovery, DOI: 10.1080/17460441.2024.2367014

To link to this article: https://doi.org/10.1080/17460441.2024.2367014

Published online: 18 Jun 2024.

EXPERT OPINION ON DRUG DISCOVERY
https://doi.org/10.1080/17460441.2024.2367014

Taylor & Francis
Taylor & Francis Group

REVIEW

Artificial intelligence for small molecule anticancer drug discovery

Lihui Duo[a], Yu Liu[a], Jianfeng Ren[a], Bencan Tang[a] and Jonathan D. Hirst[b]

[a]Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo, China; [b]School of Chemistry, University of Nottingham University Park, Nottingham, UK

ABSTRACT

Introduction: The transition from conventional cytotoxic chemotherapy to targeted cancer therapy with small-molecule anticancer drugs has enhanced treatment outcomes. This approach, which now dominates cancer treatment, has its advantages. Despite the regulatory approval of several targeted molecules for clinical use, challenges such as low response rates and drug resistance still persist. Conventional drug discovery methods are costly and time-consuming, necessitating more efficient approaches. The rise of artificial intelligence (AI) and access to large-scale datasets have revolutionized the field of small-molecule cancer drug discovery. Machine learning (ML), particularly deep learning (DL) techniques, enables the rapid identification and development of novel anticancer agents by analyzing vast amounts of genomic, proteomic, and imaging data to uncover hidden patterns and relationships.

Area covered: In this review, the authors explore the important landmarks in the history of AI-driven drug discovery. They also highlight various applications in small-molecule cancer drug discovery, outline the challenges faced, and provide insights for future research.

Expert opinion: The advent of big data has allowed AI to penetrate and enable innovations in almost every stage of medicine discovery, transforming the landscape of oncology research through the development of state-of-the-art algorithms and models. Despite challenges in data quality, model interpretability, and technical limitations, advancements promise breakthroughs in personalized and precision oncology, revolutionizing future cancer management.

ARTICLE HISTORY
Received 22 April 2024
Accepted 7 June 2024

KEYWORDS
Drug discovery; machine learning; artificial intelligence; cancer; small molecules

1. Introduction

The transition from conventional cytotoxic chemotherapy to the exploration and development of small-molecule cancer drugs has led to a rising number of successful therapies, positively impacting the lives of a significant population of cancer patients [1]. Because of this, small-molecule anticancer therapy is predominant in cancer treatment, driven by advantages such as pharmacokinetic (PK) properties, patient compliance, cost, cross-country drug transportation, and ease of storage, which set them apart from macromolecule therapy [2]. Notably, their ability to penetrate cellular membranes and reach cancer targets enables them to inhibit tumor cell proliferation [1]. A total of 89 small-molecule drugs have gained endorsement from the US Food and Drug Administration (FDA) and/or National Medical Products Administration for treating various types of cancer, significantly improving patient outcomes and survival rates [2]. Examples include the epidermal growth factor receptor (EGFR) inhibitors gefitinib and erlotinib in treating non-small cell lung cancer, erb-B2 receptor tyrosine kinase 2 (ERBB2) inhibitor lapatinib for curing ERBB2-positive breast cancer, the kinase inhibitor sorafenib for targeting the vascular epidermal growth factor receptor [3], and the kinase B-Raf Proto-Oncogene (BRAF) inhibitor vemurafenib for the treatment of metastatic melanoma with the BRAF V600E mutation [1] (Figure 1). Concurrently, thousands of chemical compounds are undergoing clinical trials for cancer therapy, with a substantial proportion of these showing promise and progressing to phase III trials [2]. While significant advancements have been made, targeted small-molecule anticancer drugs face numerous challenges. These include low response rates in certain cancer subtypes, rapid development of drug resistance due to tumor heterogeneity and genetic mutations, off-target effects leading to toxicity, and limited efficiency in solid tumors. Thus, there is a need for combination therapies, personalized treatment plans, and the development of next-generation inhibitors to overcome these obstacles [2].

The drug discovery process, marked by its prolonged cycle, substantial costs, and limited efficiency, is a longstanding focal point within the industry. On average, developing an innovative drug – from research and development to marketing – entails an investment of approximately $200 million, spans 10 to 15 years with less than 5% of compounds progressing into clinical trials successfully [4,5]. Such an arduous pipeline underscores the imperative for substantial advancements to accelerate development cycles and mitigate expense. While technologies such as high-throughput screening (HTS), network pharmacology, and RNA sequencing have offered some improvements, their impact on accelerating drug discovery remains marginal. Hence, innovative solutions are stilled awaited.

With the rise of artificial intelligence (AI) and the increased access to large-scale datasets, machine learning

CONTACT Jianfeng Ren ✉ Jianfeng.Ren@nottingham.edu.cn; Bencan Tang ✉ bencan.tang@nottingham.edu.cn Faculty of Science and Engineering, University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo 315100, China; Jonathan D. Hirst ✉ jonathan.hirst@nottingham.ac.uk School of Chemistry, University of Nottingham University Park, Nottingham NG7 2RD, UK

Article highlights

- Traditional drug discovery faces challenges in analyzing extensive datasets and unveiling hidden patterns efficiently, but AI and ML, especially DL, are revolutionizing this process by leveraging advanced computational techniques to accelerate the development of new therapeutic agents.
- Key milestones in AI-driven drug discovery, such as GENTRL and AlphaFold3, illustrate advancements in clinical development and precision in drug discovery, though challenges like data bias and the need for high-quality datasets remain.
- AI applications, leveraging techniques like classification, neural networks, and clustering, integrate complex multi-omics data to identify novel and reliable therapeutic targets, enhancing the understanding of carcinogenesis.
- AI-driven generative models, including RNNs, GANs, and VAEs, autonomously learn molecular characteristics and create innovative compounds, enabling rapid exploration of wider chemical space.
- AI-driven techniques in pharmacological prediction, preclinical applications, and drug repurposing are crucial for predicting DTIs, physicochemical properties, and ADMET characteristics, and discovering novel therapeutic strategies, thereby streamlining the drug development process.The future of AI-driven cancer discovery holds promise for revolutionizing oncology management through the integration of AI tools in every step of drug discovery, personalized medicine, and prognostic prediction, ultimately democratizing drug development and improving patient outcomes.

(ML), particularly deep learning (DL), is emerging as a paradigm-shifting force poised to reshape the drug discovery landscape [6]. By leveraging advanced algorithms, AI can go beyond conventional methodologies, utilizing vast datasets spanning genomics, proteomics, and imaging modalities [7]. This computational prowess can reveal latent patterns and relationships that might elude human perception and augment our understanding of molecular interactions, PK, and disease mechanisms, thereby presenting opportunities to expedite the discovery and optimization of novel therapeutic agents [7], especially small molecules. The conventional drug discovery pipeline involves extensive laboratory work, iterative testing, and significant resource allocation. AI, however, optimizes resource utilization by predicting the most promising candidates early in the process, thereby minimizing the need for costly, time-consuming, and labor-intensive experiments. Such acceleration extends beyond the discovery phase, as AI could streamline subsequent stages of preclinical testing and clinical trial design by identifying suitable patient populations more effectively, thereby shortening the overall development cycle [6]. Crucially, the integration of AI technologies across academia and industry has yielded tangible outcomes, as evidenced by the emergence of promising candidates swiftly advancing into clinical trials. Noteworthy examples include the serotonin receptor DSP-0038, the cyclin-dependent kinase (CDK) 7 inhibitor GTAEXS617, and the tyrosine kinase 2 inhibitor NDI-034858/TAK-279 [8]. These examples underscore the potential of AI to catalyze pharmaceutical research and to provide a competitive edge in the race to bring novel therapeutics to market.

In the following review, we trace the evolution of AI in drug discovery, emphasize various applications of AI in small molecule cancer therapeutics, outline the challenges faced, and provide some future perspectives. By doing so, we aim to enrich the understanding of the current state of AI in the small molecules anticancer drug discovery field and identify avenues for further research and innovation.

2. Historical landmarks in AI-driven drug discovery

ML and drug discovery have been closely intertwined for several decades, and there were repeated cycles of excitement and disillusionment surrounding the idea of AI drug development [9]. In the past several years, major scientific and technical discoveries have resulted in a seismic shift toward embracing computational techniques as a primary driving force for drug development in both academia and business. There were some notable successes along the way [10] and the field is still emerging. Below, we give a brief introduction to some of the historical landmarks in AI-driven drug discovery, highlighting some of the key milestones that have shaped this area (Figure 2).

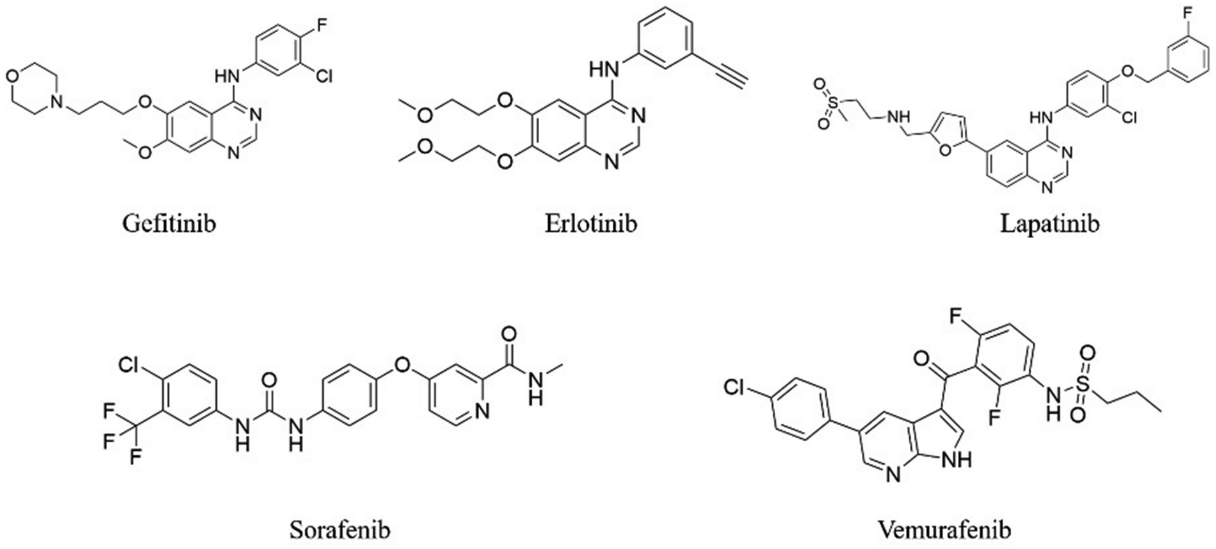

Figure 1. Some representative FDA-approved anticancer drugs.

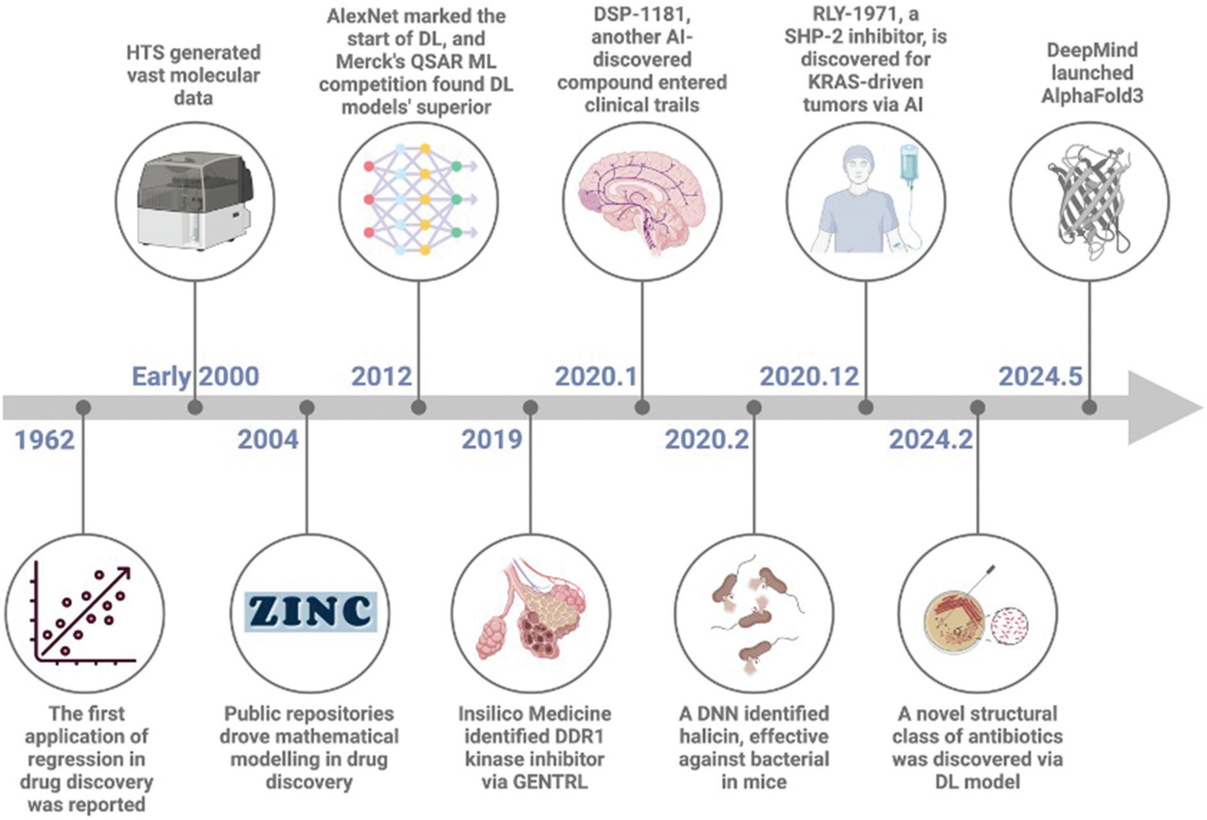
HTS generated vast molecular data
AlexNet marked the start of DL, and Merck's QSAR ML competition found DL models' superior
DSP-1181, another AI-discovered compound entered clinical trails
RLY-1971, a SHP-2 inhibitor, is discovered for KRAS-driven tumors via AI
DeepMind launched AlphaFold3
Early 2000
2012
2020.1
2020.12
2024.5
1962
2004
2019
2020.2
2024.2
ZINC
The first application of regression in drug discovery was reported
Public repositories drove mathematical modelling in drug discovery
Insilico Medicine identified DDR1 kinase inhibitor via GENTRL
A DNN identified halicin, effective against bacterial in mice
A novel structural class of antibiotics was discovered via DL model

Figure 2. Key historical milestones in AI-driven drug discovery. These landmarks (and the associated references) are discussed in the main text. Created with BioRender.com (https://www.biorender.com/).

In 1962 [11], Hansch reported the first application of regression in drug discovery to predict the properties of drugs. This early application demonstrated the potential of ML in drug discovery and paved the way for future research. Since then, the production of quantitative models and statistical methods for complicated chemical, biological, or physical phenomena started to find their place in the computational chemist's toolbox. The emergence of HTS technologies in the early 2000s allowed researchers to test large numbers of compounds rapidly for potential activity. This biotechnological advancement not only accelerated the initial phases of drug discovery but generated vast amounts of molecular data, facilitating the release of several public repositories like ZINC in 2004 [12]. These databases have promoted the application of various mathematical models in the small molecule drug development pipeline. Subsequently, ML algorithms started to play a role in helping to identify promising hits from screening libraries, enabling researchers to focus their efforts on the most promising compounds. For instance, Random Forest (RF) started to be exploited for quantitative structure-activity relationship (QSAR) and virtual screening (VS) studies [13].

In 2012, AlexNet heralded the start of the age of DL [14]. The potential for these tools to provide generalizable discoveries is strongly supported by the capacity of DL approaches to find intricate correlations between molecular representations and observations (*e.g*., toxicity, bioactivity). A QSAR ML competition, sponsored by Merck in 2012, made an early attempt to apply DL for discovering novel compounds in the pharmaceutical industry. DL models outperformed conventional ML techniques, particularly in predicting drug candidates' absorption, distribution, metabolism, excretion, and toxicity (ADMET) [15], and similar success has also been seen in toxicity prediction in the NIH Tox 21 challenge [6]. Subsequently, AI technology has shown great potential in the field of innovative drug generation, and the expansion of new indications by providing more efficient, accurate, innovative, and personalized solutions. In 2019, researchers from Insilico Medicine employed the generative tensorial reinforcement learning (GENTRL) system and identified a first-in-class anti-fibrotic small-molecule discoidin domain receptor 1 (DDR1) kinase inhibitor (INS018_055) for idiopathic pulmonary fibrosis. It was the first fully generative AI compound to enter phase II clinical trials, indicating a new level in therapeutic asset discovery for the pharmaceutical industry [16]. In 2020, Exscientia further demonstrated the potential of AI-driven small molecules by progressing the drug candidate DSP-1181, another AI-discovered compound as a treatment for obsessive-compulsive disorder, into clinical trials [17]. In the same year, Stokes *et al*. employed a DL model to predict antibacterial activity by screening more than 100 million chemical compounds, and halicin, a novel super-powerful antibiotic, was discovered to kill 35 of the world's most pathogenic bacteria even multi-drug-resistant strains, which was confirmed experimentally [18]. Similar research was also conducted recently by Wong *et al*. to identify a novel structural class that exhibited selectivity against methicillinresistant *S. aureus*, offering insights into selective antibiotic activity through interpretable ML models in antibiotic discovery [19]. These milestones underscore the accelerating momentum and critical achievements in the area of AI-driven drug discovery.

Anticancer candidates developed by companies employing AI tools in conjunction with other computational screening methods are progressing through clinical development. Relay Therapeutics, specializing in discerning candidate agents based on insights into protein dynamics, is advancing its src homology 2-containing protein tyrosine phosphatase 2 inhibitor RLY-1971 [20]. This inhibitor, developed in collaboration with Genentech through a partnership initiated in December 2020, is currently undergoing phase I clinical trials for the treatment of Kirsten rat sarcoma viral oncogene homolog-driven cancer. The prevailing approach in personalized small molecule anticancer drug discovery is target-based design, predominantly focusing on proteins as the key targets.

Very recently, Google's DeepMind launched AlphaFold3, an advanced protein structure prediction AI utilizing the diffusion-based architecture [21]. This represents a significant breakthrough, achieving unprecedented accuracy in predicting the structures and interactions of almost all biomolecules, surpassing the accuracy of existing docking tools [21]. Indeed, protein structure prediction technology has facilitated cancer target identification and hit compound recognition, culminating in the discovery of a novel and efficient hepatocellular carcinoma (HCC) CDK20 inhibitor (ISM042-2-048), as reported by Ren *et al.* [22]. This landmark signifies an advancement in our modeling of dynamic biological systems and marks a substantial progression in AI-assisted drug discovery.

Indications of the impact of AI on small-molecule drug discovery thus far primarily revolve around heightened efficiency and accelerated timelines during early-stage research. Noteworthy instances include the emergence of drug candidates featuring novel chemical structures for prominent targets or those directed at previously unexplored biological pathways. As the cohort of these drug candidates expands and advances through clinical stages in the coming years, it will become increasingly evident to what extent AI can realize its broader potential in augmenting clinical success rates and mitigating the costs associated with small-molecule anticancer therapeutics research and development. This progression will shed light on AI's capacity to usher in more innovative options for novel anticancer treatments.

3. Applications

By substantially reducing the duration and expense of new drug research, AI offers advantages across various stages of development, as shown in Figure 3. The algorithms that are reviewed in this study comprise a range of widely used approaches, advanced models and platforms that have been employed in research on small-molecule anticancer medicines.

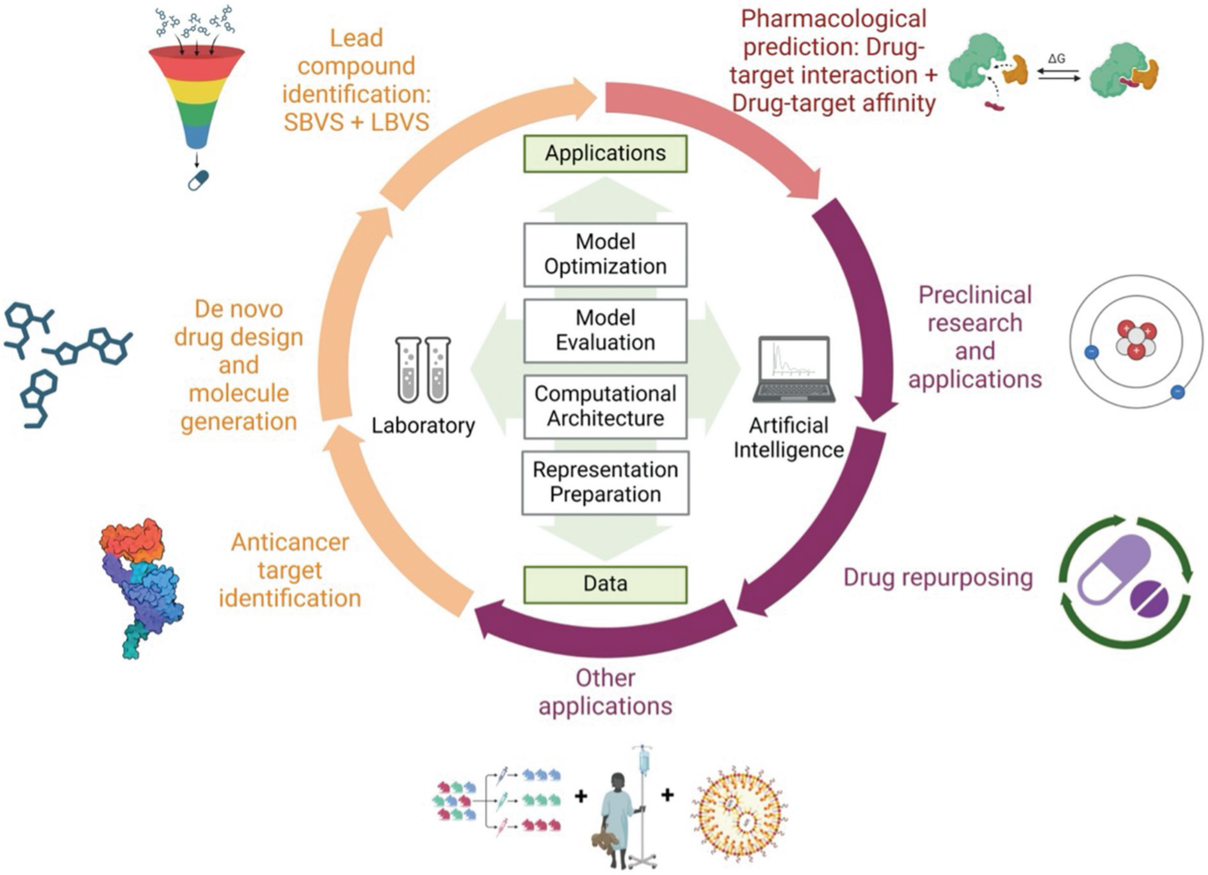
Lead compound identification: SBVS + LBVS
Pharmacological prediction: Drug-target interaction + Drug-target affinity
ΔG
Applications
Model Optimization
Model Evaluation
Computational Architecture
Representation Preparation
Data
Laboratory
Artificial Intelligence
De novo drug design and molecule generation
Preclinical research and applications
Anticancer target identification
Drug repurposing
Other applications
+
+

Figure 3. Overview of AI applications in small molecule cancer drug discovery. Created with BioRender.com (https://www.biorender.com/).

3.1. Anticancer target identification

Current targeted therapies are restricted by a limited number of exploitable targets, while the growing availability of medical omics data, spanning genomic, proteomic, transcriptomic, and epigenomic data, has opened up new opportunities for AI applications in the prediction of novel biological targets, contributing to our understanding of carcinogenesis [23].

These applications integrate enormous and complex datasets from multiple sources to identify reliable prospective targets for therapeutic medicines against human diseases [24] using techniques like classification [25], neural networks [26,27], and clustering [28]. Indeed, advancements in classification-based applications and molecular profiling techniques have enabled the utilization of genome-wide gene transcription profiles, mutational landscapes, and protein expression profiles to achieve accurate tumor subtype classification, uncovering biomarkers specific to particular tumor types. For example, pancreatic cancer subgroups and associated genetic properties were found by using classification analysis on biology networks, since such analysis can discern crucial targets by elucidating the pivotal components inside classes [29]. Jeon *et al*. constructed a support vector machine (SVM) classifier to process a multi-dimensional feature set, including mRNA expression, DNA copy number, protein-protein interaction data, and mutation prevalence, which allowed researchers to discover and prioritize potential drug targets specific to the pancreatic, breast, and ovary cancers [30]. Possible therapeutic targets for liver cancer were identified using another effective one-class SVM, where Tong *et al*. combined clinical data with gene expression patterns and protein-protein interaction networks as model inputs [31]. More recently, through the implementation of a DL-based neural network, CDK20 has been identified as a therapeutic target for HCC, and the generative AI also enabled the design of a potent small-molecule inhibitor, which exhibits specific antiproliferation effects in the corresponding cell lines [22]. Moreover, evidence suggests that the application of graph-based neural networks in cancer target identification, where molecules are represented as graphs and nodes exchange information via message passing, significantly improves target discovery accuracy by exploiting intricate network correlations and inter-node communication [23]. Wang *et al*. introduced Multi-Omics Graph cOnvolutional Networks (MOGONET), which is the first to effectively integrate multi-omics in biomedical data classification tasks by utilizing both graph convolution networks (GCNs) and cross-omics interactions for biomedical classification, resulting in the recognition of top-ranking biomarkers for targeted breast cancer therapy [26]. In a related study, Xuan *et al*. proposed the GCNLDA, based on GCNs and convolutional neural networks (CNNs), for inferring disease-related lncRNA candidates [27]. By constructing a network of lncRNA, disease, and miRNA nodes, and developing an embedding matrix for lncRNA-disease pairs, GCNLDA outperforms state-of-the-art methods through the exploration of lncRNA-disease connections within the embedding space, which has been confirmed by case studies on osteosarcoma, lung cancer, and stomach cancer in discovering potential lncRNA-disease associations. Other computational methods are also employed for predicting novel biological cancerous targets for small compounds, such as a stochastic semi-supervised ML framework DrugnomeAI [32], and a multi-model neural network PT-Finder [33]. The burgeoning integration of ML in this field signifies a transformative leap toward expeditious, cost-effective, and targeted-driven therapeutic advancements in cancer treatment modalities. However, it is essential to recognize that AI is not a standalone solution. Various platforms like SwissTargetPrediction based on ligands' similarity measurements [34], are vital complements to target identification in cancer research or serve as benchmarks for evaluating AI-driven methods. By providing additional perspectives and validation of predicted targets, these platforms with rich datasets and algorithms help to elevate the reliability and efficacy of AI-driven methods. Moreover, the complexity of the relevant biological systems often exceeds the current modeling capabilities of AI, resulting in oversimplified interpretations that might miss critical nuances. Inconsistencies in model predictions caused by the biased and heterogeneous input data can impact AI-generated outcomes. Thus, validation is important, using established biological research techniques such as imaging, microscopy, and cellular techniques for studying the phenotypic challenges within living systems.

3.2. De novo drug design and molecule generation

De novo drug design involves the creation of new small molecules without relying on existing templates. Driven by data, AI/ML-based generative models can bypass the shortcomings of standard empirical-based design paradigms, and often have much better scalability. These algorithms enable computers to autonomously learn characteristics of small molecules and suggest compounds that meet specified requirements, providing opportunities to explore vast chemistry spaces, and transforming the drug design process [35].

Recurrent neural networks (RNNs), a type of ANNs designed for sequence data, can learn molecular information from extensive sets of input simplified molecular-input line-entry system (SMILES) strings and generate compounds that exhibit similar activities to those in the training set templates while introducing innovative molecular scaffolds [36]. Despite their auspicious emergence, RNNs have encountered several hurdles, most notably the proportion of valid SMILES, intrinsic diversity, and high similarity between created molecules and compounds in the training set. To overcome these obstacles, reinforcement learning (RL) has been extensively combined with various generators to devise deep generative models for exploring broader chemical spaces and automatically designing novel molecular scaffolds [37]. In 2018, Popova *et al*. proposed RL for Structural Evolution (ReLeaSE) by training generative and predictive RNNs separately and combining them with RL settings to generate chemically feasible SMILES strings for janus protein kinase 2 inhibitors with desired properties [38]. However, such string generation has been questioned, due to concerns about SMILES degeneracy and the limited expressiveness of essential chemical properties [39]. With the rise of graph neural networks (GNNs), a series of graph convolution frames have demonstrated their superiority over SMILES strings on molecular generation [39,40], like multi-objective strategy DeepGraphMolGen [41] and the Graph Convolutional Policy Network (GCPN) [42].

Apart from the relatively conventional generative techniques mentioned above, generative adversarial networks (GANs), which consist of two neural networks contesting with each other to generate new data, and autoencoders (AEs), which are designed to learn efficient coding of input data, are widely adopted architectures in de novo drug design. For example, based on RL and GANs, Putin *et al*. conducted the reinforced adversarial neural computer (RANC) for the design of small organic molecules, where RANC produced distinctive structures that match predefined chemical descriptors while maintaining structural integrity [43]. Subsequently, a plethora of research has augmented the efficacy of the GANs framework, leading to the development of several variations, such as Objective-Reinforced Generative Adversarial Networks (ORGAN) [44], Objective-Reinforced Generative Adversarial Network for Inverse-Design Chemistry (ORGANIC) [45], molecular GAN (MolGAN) [46], and Adversarial Threshold Neural Computer (ATNC) [47]. Furthermore, ChemVAE, is the first DL-based variational autoencoder (VAE) to generate optimized drug-like molecules [48]. Similarly, Conditional Variational Autoencoder (CVAE) could generate compounds with certain properties including topological polar surface area, partition coefficient (log *P*), and molecular weight [49]. A hybrid VAE model was devised to construct candidates with predicted potent anticancer activity, and the generated compounds displayed strong inhibitory effects against targeted diseases [50]. Based on insight from generative pre-training (GPT) models, Bagal *et al*. (2021) used a masked self-attention transformer architecture to develop MolGPT, a novel encoder-decoder framework for molecular generation endeavors. It performed comparably to other contemporary ML frameworks for generating valid, unique, and novel molecules with desired scaffolds [51]. A more sophisticated technique is the adversarial autoencoder (AAE), which combines the AEs and GANs components to generate realistic data samples. Specifically, an AEE, comprising a generator and discriminator, was trained on molecules with anti-tumor growth activity, and the generated model produced molecules with desired properties in fingerprints (FP) form, resembling potent oncological drugs [52]. Other advanced architectures include the drug-generative adversarial network (druGAN) [53] and Latent GAN [54], with LatentGAN, proposed by Prykhodko, aiming to tailor inhibitors for specific anticancer protein targets, like EGFR, the serotonin 1A receptor, and sphingosine-1-phosphate receptor 1 [54].

One well-known generative design that is thought to outperform both VAE or GANs is the GENTRL model created by Insilico Medicine in the context of de novo DDR1 inhibitor design for fibrosis [16]. The entire process of discovery, from molecular generation by GENTRL to biochemical, and cell-based assays, as well as PK studies in mice, spans just 21 days [16]. The efficiency of AI-driven molecular generation offers a marked improvement over conventional drug discovery timelines, highlighting the transformative potential of AI in expediting and de-risking the drug development process. The development of AI molecular generation platforms has progressed extensively. For example, Li *et al*. employed Chemistry 42, a renowned platform in this domain, to produce promising small molecule inhibitors targeting the putative oncogene CDK8 for controlling acute myeloid leukemia and advanced solid tumors. The best compound had subnanomolar enzyme inhibitory activity ($IC_{50} = 0.4$ nM) and notable anti-proliferative effects ($IC_{50} = 2.4$ nM) [55]. The above instances illustrate that data-driven AI molecular generation methods are capable of producing compounds with innovative structures, contributing to the exploration of novel drug scaffolds (see Table 1). However, generative models are stochastic, and the resulting molecules exhibit significant structural variability and uneven quality.

3.3. Lead compound identification: VS

Screening is the process of evaluating, either experimentally or computationally, a large number of molecules before choosing candidates for further development. Because screening pharmaceuticals by biological experiments is inefficient, alternative computational VS methods that integrate AI/ML have introduced a fresh dynamism to this field and emerged to explore commercial chemical or in-house compound collections. VS consists of two methodologies as discussed below.

3.3.1. Structure-based virtual screening (SBVS)

SBVS, namely VS based on molecular docking, simulates the interaction between the 3D structure of receptor biomacromolecules and small compounds from databases to assess their affinity and identify prospective ligands by ranking their therapeutic potential [56]. Previously, molecular docking has dominated the field, utilizing programs such as AutoDock Vina, SwissDock/EADock, and Glide [57]. However, the approximate scoring functions, with largely empirically based parameters, can lead to inaccurate prediction of binding affinities and result in false positives and negatives [57]. These constraints can pose challenges in reliably identifying the most promising therapeutic candidates, and ultimately compromise the efficacy of the VS process. Ballester and coworkers initiated the application of ML regression techniques to develop AI-based scoring methods, markedly improving the predictive accuracy of SBVS [58]. This innovative strategy has been extensively adopted in VS and is a substantial contribution to enhancing the performance of the scoring function [59]. Wójcikowski *et al*. compared various docking tools and scoring functions, concluding that RF-Score-VS, an ML-based method, outperformed conventional scoring functions [60]. Other AI-based scoring functions have also been created to improve the accuracy of identifying active ligands. Notable examples include SVM-based ID-Score [61] and artificial neural network (ANN)-based NNScore [62]. In oncological research, these strategies can be useful in addressing the wide range of neoplastic disorders characterized by hallmarks of cancer. According to Wijewardhane *et al*., an ML-based scoring function is ideal for identifying small chemical inhibitors of the PD-1/PD-L1 interaction [63], which is an exciting cancer target. A small molecule drug would represent a revolution in

Table 1. Summary of molecular generative models.

Model (reference)	Description
RNN-based	
ReLeaSE [36]	□ integrates generative RNNs and predictive deep neural networks (DNNs) by training them separately and then jointly. □ creates chemically feasible molecules with desired properties.
Graph-based	
DeepGraphMolGen [39]	□ focuses on the generation of novel molecules with tailored interaction properties through a multi-objective optimization framework. □ learns interaction binding models from empirical data and incorporates a robust loss function to handle potential property score errors with GCN.
GCPN [40]	□ based on GCNs and designed for goal-directed graph generators using RL.
GAN-based	
RANC [41]	□ designed for de novo design of small-molecule organic structures using GANs and RL. □ is introduced as a generative model for small molecular graphs, eliminating the need for complex graph-matching procedures. □ employs a differentiable neural computer with explicit memory banks, which overcomes common challenges in adversarial settings.
ORGAN [54]	□ introduces a method to enhance the quality of generated samples in sequence-based generative models by combining adversarial training with expert-based rewards using RL.
ORGANIC [43]	□ optimizes the chemical space by combining GANs for creating unique molecules and RL to bias the distribution toward specific characteristics.
MolGAN [55]	□ combines GANs with RL to produce molecules with specific desired properties. □ is introduced as a generative model for small molecular graphs, eliminating the need for complex graph-matching procedures.
ATNC [45]	□ combines GANs architecture and RL, using a differentiable neural computer as a generator with an adversarial threshold block. □ an objective reward function called internal diversity clustering is employed to enhance diversity. □ trained on SMILES representations of molecules using four objective functions. □ outperforms ORGANIC, generating 72% valid and 77% unique SMILES strings with better drug-likeness properties.
AE-based	
ChemVAE [46]	□ utilizes an encoder to transform discrete molecular structures into continuous vectors, a decoder for reverse operation, and a predictor to estimate chemical properties based on the continuous vector representation. □ employs continuous representations to facilitate the generation of new chemical structures through various operations in the latent space. □ applies gradient-based optimization for efficiently seeking optimized functional compounds.
CVAE [56]	□ introduces a molecular generative model utilizing a CVAE. □ is designed to control multiple molecular properties concurrently by incorporating them into a latent space. □ demonstrates the generation of drug-like molecules processing five specific target properties as proof of concept.
GPT-based	
MolGPT [49]	□ trained on the next token prediction task using masked self-attention. □ performs comparably with other ML frameworks in generating valid, unique, and novel molecules.
AAE-based	
druGAN [51]	□ focuses on using a deep generative AAE for identifying molecular FP with predefined anticancer properties.
LatentGAN [52]	□ combines an AE and a GAN for de novo molecular design. □ is successfully applied to generate both random drug-like compounds and target-biased compounds.
GENTRAL [15]	□ was used to discover potent inhibitors of DDR1, a kinase target associated with fibrosis and other diseases, and it optimizes synthetic feasibility, novelty, and biological activity.

comparison to existing antibody therapies. On a serine/threonine kinase target, a support vector regression-based scoring function was more accurate than five classical scoring functions, enabling the discovery of low-nanomolar inhibitors [64]. Similarly, on the estrogen receptor alpha target, ANN-based scoring functions identified hits that were both chemically innovative and biologically active *in vitro*, with the best achieving mid-nanomolar potency [65]. The integration of a scoring function into a GNN-based model enabled effective gene analysis, providing insights into the associated drug discovery [66]. While these approaches demonstrate notable efficiency and have the potential to mitigate false-positive rates in SBVS, ML-based scoring functions still struggle to make accurate predictions on compounds that differ substantially from the structures in the training set [67]. Furthermore, efficiency in structure-based lead compound discovery could also be improved by combining ML models with molecular docking-based VS. Xie *et al.* coupled a two-stage SVM with docking-based approaches to find novel c-Met kinase inhibitors for thyroid cancer, yielding active hits from a pool of 18 million molecules [68]. Valarmathi *et al.* found that the aforementioned c-Met kinase could be inhibited by 1-amino-5-chloro-anthraquinone based on molecular docking [69]. Several SBVS studies have identified inhibitors targeting other critical targets implicated in cancer development, including CDK4/6 [70], phosphoinositide 3-kinase protein [71], histone deacetylases [72], and vascular endothelial growth factor A [73].

3.3.2. *Ligand-based virtual screening (LBVS)*

In contrast to the SBVS strategy, which is constrained by the requirement for structural data of the target protein, LBVS utilizes knowledge of the bioactivity of known ligands to screen possible lead compounds from a vast library of compounds. According to the premise that compounds with comparable structural properties would have comparable biological activity [74], VS based on ligands can be categorized into pharmacophore modeling, two-dimensional (2D)/3D structural similarity searching, and ML-based LBVS. Among these, ML-based LBVS has emerged as an attractive strategy. By establishing robust relationships between molecular features and assay results, it employs a regression model for compound activity prediction and classification models to categorize compounds based on their structural similarity. Some popular QSAR-based LBVS include RF, SVM, Bayesian algorithm, and more advanced models like multitask DL models [6]. Valentini *et al.*, for example, utilized conventional ML-guided VS approaches, including logistic regression, k-Nearest Neighbour (kNN), gradient boosting, RF, and SVM, in the realm of cancer therapeutics exploration [75]. Along with *in vitro* and *in vivo* experiments, they identify two novel potential anti-apoptotic protein inhibitors, which exhibited excellent binding to recombinant myeloid cell leukemia-1, B-cell lymphoma-extra-large, and B-cell lymphoma-2 proteins, and demonstrated potent efficacy against diverse tumor histocytes [75]. Similarly, two novel CDK5 inhibitors (CPD1 and CPD4) for ovarian and colon cancer were identified through the utilization of QSAR-guided LBVS approaches, including RF, SVM, kNN, and multilayer perceptron (MLP) [76]. Beyond the traditional ML models, the use of DL, incorporating various architectures like GNNs, has expanded the potential and benefits of LBVS, showing exceptional predictive capabilities, reliable feature extraction, and minimal generalization error. More specifically, Zhang *et al.* focused on the widely studied anticancer target, the CDK series. They developed ten classification models employing two conventional ML and four DL methodologies to distinguish between CDK9 inhibitors and non-inhibitors. These models, constructed based on molecular FP and graphs, highlighted the FP-based GNN's predictive accuracy, and their findings suggested one potential inhibitor for curing leukemia [77]. Another DL classification algorithm (DNN-VS) with the Spark-H2O platform was developed to improve the bioactivity prediction, and results on the protein tyrosine kinase and receptor for breast cancer indicated that the model exhibited state-of-the-art performance [78]. Xu *et al.* employed CNNs to directly input molecular images for the screening of CDK4 inhibitors, achieving better performance than competing models [79]. Other DL approaches, such as RNNs, and RL, have been increasingly studied for LBVS in recent years, as such DL strategies show great promise for anticancer drug screening and prioritization [80]. Nevertheless, several bottlenecks persist and need to be overcome. These include: 1. Quality and utility of chemical representation, necessitating refinement in encoding methodologies to capture intricate molecular features [81]. 2. Generalization across chemical space, demanding robust models capable of extrapolating knowledge to novel chemical entities, notwithstanding structural diversity [81]. 3. Data biases induced by the intricacies of cancer, particularly the influence of tumor microenvironment, requiring meticulous curation of datasets to mitigate biases and enhance representativeness [82]. 4. Activity cliffs, where minor structural modifications lead to substantial changes in biological activity, underscoring the need for algorithms capable of discerning subtle structural variations [82].

3.4. Pharmacological prediction

3.4.1. Drug-target interaction (DTI) prediction

Accurate prediction of newly discovered DTIs, based on the amino acid sequence, plays a pivotal role in the success of lead compound discovery. It enables understanding of therapeutic efficacy, facilitates drug repurposing, and prevents polypharmacology protein interactions, ultimately enhancing cancer treatment [83]. However, the limitations of costly and laborious *in vitro* wet experiments, such as yeast two-hybrid screening [84], phage display technology [85], and HTS, have led to the widespread application of *in silico* methods in the field, as presented below.

DTI prediction is often treated as a binary classification problem: is there an interaction or not? ML-based DTI models are expected to detect more information on the mechanism of DTI and make classification assessments from encoded features of target proteins and small molecules. In 2011, Wang *et al.* employed an SVM-based model trained on 15,000 protein-ligand interactions [86]. Leveraging primary protein sequences and structural characteristics of small molecules, this *in silico* model has achieved accurate predictions for DTI, which enabled the discovery of nine novel compounds and corresponding interactions with four critical targets, including silent information regulator sirtuin 1 involved in cell survival, metabolism, and proliferation, with implications for various cancers [86]. Apart from relying solely on the encoding of amino acids and small molecules, it is essential to integrate diverse heterogeneous data sources for predicting DTI. By integrating pharmacological and chemical data, Yu *et al.* utilized two RF models to predict potential drug-protein interactions with high sensitivity and specificity, outperforming benchmarks like SVM [87]. These models not only facilitated the prediction of drug-target associations but also extended to target-disease and target-target associations. Compared to conventional ML methods, DL-based techniques frequently show greater accuracy in predicting DTI [88]. Several DL-based algorithmic frameworks, such as DeepDTIs [89], DeepConv-DTI [90], and Deep NP [91], have been developed. Among various variations, pairwise input neural networks (PINNs) with inputs represented as pairs of target-ligand feature vectors, are well-suited for the prediction of pairwise relations, especially DTI [92]. This was verified by Wang *et al.*; a five-fold cross-validation evaluation demonstrates that the proposed PINNs are better than other representative methods for predicting DTI [93]. In addition, graph-based neural networks identified of novel DTIs by leveraging integrated features to mitigate high false positive rates, thereby improving the predictive reliability of the model [94]. For example, in line with the success of GNNs models, Wen and colleagues incorporated dual-branched self-supervised pre-trained molecular graph models and protein sequence models, culminating in the construction of a transformer architecture for DTI modeling. This model demonstrated promising outcomes in predicting CDK12 interactions, revealing five previously unknown CDK12 inhibitors [95]. In the majority of cases, DTI prediction is treated as a binary classification problem, which neglects more subtle variations in target-inhibitor binding affinities.

3.4.2. Drug-target affinity (DTA) prediction

The effectiveness of drugs depends on their ability to interact with and bind to the target receptors. Compounds lacking affinity toward the intended target may fail to produce the desired therapeutic response. It is also likely that compounds in certain cases interact with proteins or receptors that were not anticipated, resulting in the possibility of adverse effects. Therefore, in the context of cancer research, DTA is a critical area of investigation. Even though measuring inhibition constants and dissociation are used to evaluate the binding affinity empirically, these experiments can be time-taking and expensive. AI regression approaches have been explored to expedite DTA. There are two main categories: sequence-based and graph-based methodologies.

Sequence-based DTA prediction entails the utilization of biological sequences, such as protein and nucleotide sequences, as input data for modeling binding affinity.

Öztürk *et al*. 2018 introduced DeepDTA, a DL model. Molecules were encoded using SMILES representations, and protein targets were encoded using amino acid sequences, which were utilized as input for a CNN model [88]. DeepDTA outperformed conventional ML-based methods like Kronecker-Regularized Least Squares and SimBoost in predicting drug-target binding affinity. Building upon the success of DeepDTA, a series of integrated DL models have been developed for predicting DTA; relying solely on single-feature representation fails to fully characterize small molecules or proteins [83]. WideDTA is another CNN-based DL method that integrates ligand SMILES, amino acid sequences, ligand-protein maximum common substructures, as well as protein domains and motifs as input features [91], whereas DeepAffinity is an interpretable DL model that incorporates both RNN and CNN architectures, utilizing both labeled and unlabeled data to consider the compounds in SMILES format, protein sequences and their physicochemical properties [83].

Graph-based methods employ graph representations of molecular and protein structures. Evidence suggests that graph-based neural networks outperform those relying solely on CNNs. This advantage stems from their ability to incorporate drug structure and DTI information, obtain interaction modes, and capture structural nuances that are often overlooked in sequence-based models [96]. Nguyen et al. were among the first researchers to utilize GNNs for DTA prediction. The neural network model that they introduced, GraphDTA, proved to be well suited for DTA regression tasks [96]. A more advanced framework, the graph early fusion affinity model, utilizes contact maps to represent structural attributes of proteins and integrates an attention mechanism to facilitate interactions between nodes associated with drug molecules and those associated with amino acids [97]. MGraphDTA adopted a multi-scale GNN to encode information regarding molecular substructures, and achieved notable performance on benchmark datasets [98]. Collectively, the transformative impact of AI on DTA has led to a step-change in cancer drug discovery paradigms, paving the way for advancements in targeted therapy. Nevertheless, the challenges in DL for DTA prediction include difficulties in protein and drug representation, feature fusion, as well as model generalization.

3.5. Preclinical applications

3.5.1. Physicochemical property prediction

Knowledge of physicochemical properties, such as log *P*, solubility, intrinsic permeability, and degree of ionization, is critical for the discovery or design of molecules with desired PK and pharmacodynamic profiles. AI technologies enhance the prediction of these properties compared to traditional manual experiments, allowing for rapid iteration and reducing reliance on specialized equipment, reagents, and labor. Such computational methods are not only affordable and scalable but also instrumental in predicting properties that are challenging to measure experimentally. Zang *et al*. utilized various ML (multiple linear regression, RF, SVM) and FP protocols to predict six physicochemical parameters simultaneously, including log *S*, log *P*, boiling point, melting point, bioconcentration factor and vapor pressure [99]. Some research in this area has concentrated on developing different models for a single property prediction, especially aqueous solubility, as it has a direct impact on the pharmacological efficacy of compounds, but accurately predicting this property remains a formidable task. Panapitiya *et al*. investigated several DL methods (including fully connected neural networks, RNNs, GNNs, and SchNet) and molecular representation approaches (including SMILES, molecular descriptors, 3D atomic coordinates, and molecular graphs) for solubility prediction [100]. The fully connected neural network demonstrated superior performance among those models, with 2D molecular descriptors notably contributing the most to the predictive accuracy. Although molecules can be represented in various ways, predictions for a particular property often depend on specific features, such as the connectivity indices of various molecules [6], and the number of hydrogen bonds [101], which are correlated with solubility. DL methods can enhance the predictive power in molecular physicochemical property prediction [15], leading to the development of a series of advanced models, such as undirected graph recursive neural networks [102], the model combined with natural language processing known as SolTranNet [103], and the architecture based on GNN – MoGAT.

Significant AI-assisted computational architectures are also dedicated to predicting another vital aspect of drug development and discovery, the permeability coefficient of small molecules. This pertains to the fundamental role of permeation across biological membranes, facilitated through passive transmembrane diffusion and/or active transport mechanisms, in numerous pharmaceutically relevant biological processes, resulting in the absorption of oral medications [104]. AI predictors primarily utilize data from *in vitro* permeability assays, a well-known technique, to train models for predicting cellular permeability. For example, the Caco-2 cell-based ML models, constructed via the AutoQSAR621 system, correlate intrinsic permeability with a molecular structure [105]. Similarly, using Caco-2 permeability data, a permeability coefficient prediction model incorporating 30 descriptors was trained on 1272 molecules, employing boosting, SVM regression, partial least-squares and multiple linear regression [106]. AI-based tools have also been utilized to model other physicochemical data like acid dissociation constants [107]. The influential role of these physicochemical properties and their interdependence set the foundation for the accurate prediction of ADMET characteristics.

3.5.2. The prediction of ADMET

Early identification of ADMET characteristics is crucial in the drug development process, as poor ADMET profiles often lead to the failure of candidates in the later stages of clinical development and market entry. The impracticality of assessing the ADMET properties of numerous compounds through lengthy animal trials has led to growing interest in the creation of *in-silico* ADMET prediction models [108].

The prediction and optimization of ADMET characteristics in the development of cancer drugs have been transformed by the advent of AI algorithms. For instance, a computational study of exemestane, a classic drug for end-stage breast carcinoma, pinpoints the exact cytochrome P450 aromatase binding location [109]. Moroy *et al.* explored the identification of potential candidates that can disrupt hormonal systems via binding into the androgen receptor (AR) for evaluation [110,111], and they accurately predicted the binding energy of 119 AR ligands and identified three antipsychotic drugs as weak AR antagonists, consistent with observed side effects in patients, demonstrating the potential for improving prostatic cancer drug discovery. In another technique, Tox(R)CNN employed a deep CNN algorithm to evaluate the cytotoxicity of medicines exposed to DAPI-stained cells [112]. This approach offers sensitive, robust, and cost-effective techniques for *in vitro* screening of cancer drug-induced toxicity, potentially diminishing the reliance on traditional *in vivo* experiments. While widely applied, many integrated learning approaches are still addressing challenges like imbalanced datasets and high dimensionality. Ensemble approaches like adaptive ensemble classification framework and multi-task graph convolutional models, may be solutions to deal with these issues, but for more complex tasks such as *in vitro* metabolic stability prediction, the latter did not achieve good results due to the simplicity of the model [80]. Several user-friendly ADMET tools and online servers have been created to predict ADMET characteristics of chemicals accurately, with SwissADME [113], ADMETlab3.0 [114], and OptADMET [115] being widely praised. Due to variations in training datasets and model parameters, certain properties predicted by different models differ from comparable tools. However, those user-friendly platforms are particularly beneficial for non-expert users, providing comprehensive and relatively accurate ADMET properties, thus serving as valuable resources for medicinal chemists. In short, driven by omics, AI-assisted ADMET predictions allow for the early selection of promising drug candidates with proper drug-likeness properties, which might bridge the gap between laboratory research and clinical applications.

3.6. Drug repurposing

Drug repurposing identifies new therapeutic uses for an 'old' medicine outside its current medical indication. The repurposing of existing drugs permits their direct qualification in phase II clinical trials, resulting in a more economical, safe, and speedy alternative to traditional drug discovery. Relaunching an existing drug requires approximately US$8 million compared to the launch of a new drug entity, which can cost around US$41 million [83].

Drug reprofiling necessitates considering multiple factors, including the implementation of multiscale models to define distinct networks specific to particular cancers, and AI techniques have facilitated concrete examples of drug repurposing by considering various heterogeneous data. Computational-driven supervised learning, employing ML approaches, like SVM, RF, and neural networks, has been widely adopted to help evaluate whether a drug could be repurposed for a novel medical indication [6]. DNNs have also been used to predict the therapeutic use category for pharmaceuticals including antineoplastic and select repurposing candidates by their chemical structural similarities with licensed cancer medications [116]. Specifically, the application of the cellular network-based DL technology, deepDTnet, has been investigated to predict the therapeutic potential of topotecan, a well-established topoisomerase inhibitor, and results demonstrated promise in repurposing topotecan for the treatment of multiple sclerosis through the inhibition of human retinoic acid receptor-related orphan receptor-gamma-t [117]. Cheng *et al.* introduced another network-based method that could accurately predict drug responses via integrating transcriptome profiles with protein-protein interactome, drug-target interactions, drug-induced microarray data, and whole-exome sequencing. They also focused on identifying novel applications for existing cancer medications, specifically targeting substantially altered genes or those genes' neighboring genes within the human protein-protein interaction networks [118,119]. With the advancement of graph-based neural networks, a growing number of related repurposing models suggested a distinct advantage in increasing the success of drug reuse. The models made use of feature information from drug-drug links and drug-cancer pairs. In particular, GraphRepur, a GNNs model proposed by Cui *et al.*, validated some effective compounds for breast cancer using test sets from the literature and this method outperformed models like the RF, DNN, and GCN [120]. In supervised models for therapeutic switching tasks, cross-validation analyses often have high performance metrics (*e.g.*, the area under the curve ranging from 0.75 to 0.95), demonstrating their ability to identify new drug-disease associations [6]. However, their practical application, particularly for rare diseases, is hindered by the prerequisite of known and similar drugs for a specific condition of interest. Additionally, the lack of high-quality negative samples in some databases results in a higher false-positive rate, which is suboptimal [6]. To address these challenges, semi-supervised and unsupervised learning methods are expected to play an increasing role in drug reprofiling, as they can overcome limitations of data availability. An example of an unsupervised approach is a robust and automatic web tool called mode of action by network analysis (MANTRA), which could predict similarities in drug effects and their mechanisms of action (MoA) [121]. This constructed drug network proved capable of not only identifying drug communities with similar MoA or shared pathways, but also correctly predicting MoA for nine anticancer compounds. In addition, this model revealed an unreported effect for a well-known drug, indicating the potential of MANTRA in MoA prediction for antitumor agents, exemplified by the correct classification of CDK2 inhibitors and their similarity to topoisomerase inhibitors. A series of notable semi-supervised learning algorithms for drug reprofiling is the laplacian regularized least-squares (LapRLS), NetLapRLS, and label propagation with mutual interaction information derived from heterogeneous networks [6].

Different computational pipelines, such as PREDICT, and SPACE, also consider a systematic analysis of big data, including the drug-drug, disease – disease similarity, the similarity between target molecules, chemical structures, and gene expression profiles while forming the hypothesis of drug repositioning [6,122]. The application of these methods has yielded successful discoveries, as evidenced by PREDICT, which identified novel therapeutic uses, like progesterone for a rare form of renal cell carcinoma, supported by existing literature [123]. Overall, the significant role of AI in drug repurposing research has led to the development of numerous successful models and predictions for identifying the repositioning of anticancer candidates.

3.7. Other applications

AI has also emerged as a transformative technology with applications across other domains in drug discovery, including designing clinical trials with AI assistance, post-market monitoring, and prognostic prediction [80]. For example, Patel and colleagues [124] pointed out that certain ML algorithms could be regarded as a substitute for animal trials to predict the bitterness of various molecules used in medicines [125]. In a healthcare setting, AI is also used to anticipate a patient's reaction to therapy and the therapeutic effectiveness of an anticancer drug [126]. Notably, recent research demonstrated the potential of DL to establish connections between histological patterns observed in whole slide images of hematoxylin and eosin stained breast cancer sections and drug sensitivities derived from cell lines [127]. By leveraging gene expression-based mapping techniques to infer drug effects on cancer cell lines, this sophisticated DL model was trained to predict patients' sensitivity to a range of drugs directly from whole slide images. Such groundbreaking methodology harnesses the power of routine whole slide images to anticipate drug sensitivity profiles for both approved and experimental drugs, while also uncovering cellular and histological patterns associated with these profiles in cancer patients. The progress of individualized medicine has also been widely addressed by AI in the recent area of big data, notably omics-based precision medicine. Indeed, Huang *et al.* focused on precision medicine and used SVM with conventional recursive feature removal to predict the individual response of each patient to medicines using gene expression patterns [128]. Furthermore, there is extensive utilization of AI in drug release and formulation, such as nanocarrier self-assembly prediction and drug-excipient combination analysis via computational methods [80]. On the whole, ML provides an excellent opportunity to evaluate chemical data and provide insights that can aid in drug development.

4. Conclusion

This review highlights the historical achievements in AI-driven drug discovery and emphasizes the broad range of applications of AI in small-molecule cancer drug discovery, from target identification, de novo drug design, validation of lead identification, and pharmacological prediction to preclinical and drug repurposing studies. We hope that the insights provided in this review could contribute to a deeper appreciation of the current state of AI in the field and the challenges still faced. By addressing these hurdles, AI holds immense potential to mitigate the global burden of cancer and improve patient outcomes through the development of novel small-molecule targeted therapies.

5. Expert opinion

The transitions from omics-based biomedical data to multi-dimensional big data have brought transformations across all stages of drug development through the widespread adoption of AI technology. Advanced AI algorithms are impacting real-world cancer therapeutic outcomes by accelerating the identification of potential drug targets, designing novel therapeutic agents, and optimizing treatment regimens for cancer patients. However, the implementation of AI-driven approaches within complex, multifaceted drug development processes encounters several challenges, including issues related to data quantity and quality, model interpretability and validation, as well as technological limitations [129]. Ensuring the accuracy and reliability of AI models necessitates the curation of larger volumes of high-quality datasets, the enhancement of data standardization and sharing practices, and the development of robust validation frameworks [130]. Additionally, advancements in post hoc explanation techniques and trusted AI methods can enhance model interpretability without compromising accuracy and alleviate 'black box' issues. Addressing technical challenges such as poor generalization, usability, and repeatability of models requires ongoing research and development efforts focused on data augmentation, transfer learning, ensemble models, code sharing, and fostering interdisciplinary collaboration [80]. The efforts made in AI-driven methodologies have the potential to set new standards in oncology research, paving the way for breakthroughs, and establishing a new paradigm in the fight against cancer.

Thinking prospectively, the future of AI-driven cancer drug discovery is compelling, with ongoing advancements expected to shape oncology management in the coming years. Standard procedures in drug discovery will likely incorporate AI-driven tools and techniques in every step, revolutionizing the way new therapies are identified, validated, and brought to market. Advances in technology would enable the collection of vast amounts of data at an individual patient level. Such wealth of data information could further empower data-driven AI to enhance decision-making precision, tailoring highly personalized cancer prevention and treatment plans. While the current focus is on small-molecule cancer drug discovery, there are also promising opportunities in other therapeutic areas and modalities, including biologics, gene therapy, and regenerative medicine. For example, one emerging implication is to integrate AI with clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 based gene-editing technology to revolutionize precision oncology [131]. By optimizing guide RNA design, predicting off-target effects, and identifying effective gene editing strategies, AI could enhance the precision and efficiency of CRISPR, thereby accelerating the development of gene therapies tailored to individual genetic profiles for accurate cancer cell targeting with improved efficacy and safety profiles. Such synergy could also facilitate the discovery of novel genetic targets, and innovative treatments for currently untreatable cancers, significantly advancing precision oncology and revolutionizing future cancer management by improving patient outcomes and reducing resistance [131].

We are in the midst of a paradigm shift. The field of AI-driven drug discovery is poised to evolve rapidly. In the future, we can expect to see the widespread adoption of sophisticated AI approaches in cancer drug development, with greater emphasis on personalized medicine, predictive modeling, and precision oncology, thus democratizing this field, reducing risks, and enabling new entrants.

Abbreviation

AAE	Adversarial Autoencoder
AEs	Autoencoders
ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity
AI	Artificial Intelligence
ANN	Artificial Neural Network
AR	Androgen Receptor
ATNC	Adversarial Threshold Neural Computer
BRAF	B-Raf Proto-Oncogene
CDK	Cyclin-Dependent Kinase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CNNs	Convolutional Neural Networks
CVAE	Conditional Variational Autoencoder
DDR1	Discoidin Domain Receptor1
DTA	Drug-Target Affinity
DTI	Drug-Target Interaction
DL	Deep Learning
DNNs	Deep Neural Networks
druGAN	drug-Generative Adversarial Network
EGFR	Epidermal Growth Factor Receptor
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
FDA	Food and Drug Administration
FP	Fingerprints
GANs	Generative Adversarial Networks
GCPN	Graph Convolutional Policy Network
GCNs	Graph Convolution Networks
GENTRL	Generative Tensorial Reinforcement Learning
GNNs	Graph Neural Networks
GPT	Generative Pre-Training
HCC	Hepatocellular Carcinoma
HTS	High-Throughput Screening
kNN	k-Nearest Neighbour
LapRLS	Laplacian Regularized Least-Squares
LBVS	Ligand-Based Virtual Screening
Log P	Partition Coefficient
MANTRA	Mode of Action by Network Analysis
ML	Machine Learning
MLP	Multilayer Perceptron
MoA	Mechanisms of Action
MOGONET	Multi-Omics Graph cOnvolutional Networks
MolGAN	Molecular GAN
ORGAN	Objective-Reinforced Generative Adversarial Networks
ORGANIC	Objective-Reinforced Generative Adversarial Network for Inverse-Design Chemistry
PINNs	Pairwise Input Neural Networks
PK	Pharmacokinetic
QSAR	Quantitative Structure-Activity Relationship
RANC	Reinforced Adversarial Neural Computer
ReLeaSE	Reinforcement Learning for Structural Evolution
RF	Random Forest
RL	Reinforcement Learning
RNNs	Recurrent Neural Networks
SBVS	Structure-Based Virtual Screening
SMILES	Simplified Molecule-Input Line-Entry System
SVM	Support Vector Machine
VAE	Variational Autoencoder
VS	Virtual Screening
2D	Two-Dimensional
3D	Three-Dimensional

Funding

This work was supported by the National Natural Science Foundation of China [Grant No. 22171153], the Ningbo Science and Technology Bureau under CM2025 program [Grant No. 2020Z092], the Ningbo Natural Science Foundation Programme [Grant No. 2022J171], the Ministry of Science and Technology of the People's Republic of China under funding scheme National Key R&D Program of Intergovernmental Kay Projects [Grant No. 2018YFE0101700], and the Ningbo Municipal Key Laboratory on Clean Energy Conversion Technologies [Grant No. 2014A22010] as well as the Zhejiang Provincial Key Laboratory for Carbonaceous Waste Processing and Process Intensification Research funded by the Zhejiang Provincial Department of Science and Technology [Grant No. 2020E10018]. The authors also acknowledge The University of Nottingham, England and the University of Nottingham Ningbo China for the use of their High-Performance Computing facilities. Finally, JD Hirst acknowledges support from the Department of Science, Innovation and Technology (DSIT) and the Royal Academy of Engineering under the Chairs in Emerging Technologies scheme.

Declaration of interest

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Reviewer disclosures

One review is an employee of Charles River Laboratories. Peer reviewers on this manuscript have no other relevant financial or other relationships to disclose.

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Hoelder S, Clarke PA, Workman P. Discovery of small molecule cancer drugs: successes, challenges and opportunities. Mol Oncol. 2012;6(2):155–176. doi: 10.1016/j.molonc.2012.02.004
2. Zhong L, Li Y, Xiong L, et al. Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. Signal Transduct Target Ther. 2021;6(1):201. doi: 10.1038/s41392-021-00572-w

•• **This review discusses the landscape of small molecule targeted anti-cancer drugs, highlighting approved drugs, clinical candidates, challenges, and insights for their development.**

3. Yap TA, Workman P. Exploiting the cancer genome: strategies for the discovery and clinical development of targeted molecular therapeutics. Annu Rev Pharmacol Toxicol. 2012;52(1):549–573. doi: 10.1146/annurev-pharmtox-010611-134532
4. Vatansever S, Schlessinger A, Wacker D, et al. Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: state-of-the-arts and future directions. Med Res Rev. 2021;41(3):1427–1473. doi: 10.1002/med.21764

• **This review provides a comprehensive overview of AI and ML applications in drug discovery for central nervous system diseases, covering key procedures such as target identification, compound screening, hit/lead generation, drug response prediction, de novo drug design, and drug repurposing.**

5. Dowden H, Munro J. Trends in clinical success rates and therapeutic focus. Nat Rev Drug Discov. 2019;18(7):495–496. doi: 10.1038/d41573-019-00074-z
6. Yang X, Wang Y, Byrne R, et al. Concepts of artificial intelligence for computer-assisted drug discovery. Chem Rev. 2019;119(18):10520–10594. doi: 10.1021/acs.chemrev.8b00728

7. Deng J, Yang Z, Ojima I, et al. Artificial intelligence in drug discovery: applications and techniques. Brief Bioinform. 2022;23(1): bbab430. doi: 10.1093/bib/bbab430
8. Lv Q, Zhou F, Liu X, et al. Artificial intelligence in small molecule drug discovery from 2018 to 2023: does it really work? Bioorg Chem. 2023;141:106894. doi: 10.1016/j.bioorg.2023.106894
9. Van Drie JH. Computer-aided drug design: the next 20 years. J Comput Aided Mol Des. 2007;21(10–11):591–601. doi: 10.1007/s10822-007-9142-y
10. Talele TT, Khedkar SA, Rigby AC. Successful applications of computer aided drug discovery: moving drugs from concept to the clinic. Curr Top Med Chem. 2010;10(1):127–141. doi: 10.2174/156802610790232251
11. Hansch C, Maloney PP, Fujita T, et al. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. Nature. 1962;194(4824):178–180. doi: 10.1038/194178b0
12. Carracedo-Reboredo P, Liñares-Blanco J, Rodríguez-Fernández N, et al. A review on machine learning approaches and trends in drug discovery. Comput Struct Biotechnol J. 2021;19:4538–4558. doi: 10.1016/j.csbj.2021.08.011
13. Bruce CL, Melville JL, Pickett SD, et al. Contemporary QSAR classifiers compared. J Chem Inf Model. 2007;47(1):219–227. doi: 10.1021/ci600332j
14. Alom MZ, Taha TM, Yakopcic C, et al. The history began from alexnet: a comprehensive survey on deep learning approaches. arXiv:1803.01164. 2018.
15. Ma J, Sheridan RP, Liaw A, et al. Deep neural nets as a method for quantitative structure–activity relationships. J Chem Inf Model. 2015;55(2):263–274. doi: 10.1021/ci500747n
16. Zhavoronkov A, Ivanenkov YA, Aliper A, et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. Nat Biotechnol. 2019;37(9):1038–1040. doi: 10.1038/s41587-019-0224-x
17. Kirkpatrick P. Artificial intelligence makes a splash in small-molecule drug discovery. Biopharma Deal. 2022;2022: d43747–022–00104–00107. doi: 10.1038/d43747-022-00104-7
18. Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180(4):688–702.e13. doi: 10.1016/j.cell.2020.01.021
19. Wong F, Zheng EJ, Valeri JA, et al. Discovery of a structural class of antibiotics with explainable deep learning. Nature. 2024;626 (7997):177–185. doi: 10.1038/s41586-023-06887-8

- •• **This study introduces an explainable DL approach for identifying new antibiotic classes.**

20. Taylor AM, Williams BR, Giordanetto F, et al. Identification of GDC-1971 (RLY-1971), a SHP2 inhibitor designed for the treatment of solid tumors. J Med Chem. 2023;66(19):13384–13399. doi: 10.1021/acs.jmedchem.3c00483
21. Abramson J, Adler J, Dunger J, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 2024:1–3. doi: 10.1038/s41586-024-07487-w.

- •• **This study introduces AlphaFold 3, a diffusion-based DL model that significantly enhances the accuracy of predicting biomolecular interactions, including protein-ligand, protein-nucleic acid, and antibody-antigen complexes.**

22. Ren F, Ding X, Zheng M, et al. AlphaFold accelerates artificial intelligence powered drug discovery: efficient discovery of a novel CDK20 small molecule inhibitor. Chem Sci. 2023;14 (6):1443–1452. doi: 10.1039/D2SC05709C
23. You Y, Lai X, Pan Y, et al. Artificial intelligence in cancer target identification and drug discovery. Signal Transduct Target Ther. 2022;7(1):156. doi: 10.1038/s41392-022-00994-0
24. Jin S, Zeng X, Xia F, et al. Application of deep learning methods in biological networks. Brief Bioinform. 2021;22(2):1902–1917. doi: 10.1093/bib/bbaa043
25. Zhu Y, Shen X, Pan W. Network-based support vector machine for classification of microarray samples. BMC Bioinformatics. 2009;10 (1):1–11. doi: 10.1186/1471-2105-10-S1-S21
26. Wang T, Shao W, Huang Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nat Commun. 2021;12 (1):3445. doi: 10.1038/s41467-021-23774-w
27. Xuan P, Pan S, Zhang T, et al. Graph convolutional network and convolutional neural network based method for predicting lncrna-disease associations. Cells. 2019;8(9):1012. doi: 10.3390/cells8091012
28. Sanchez R, Mackenzie SA. Integrative network analysis of differentially methylated and expressed genes for biomarker identification in leukemia. Sci Rep. 2020;10(1):2123. doi: 10.1038/s41598-020-58123-2
29. Sinkala M, Mulder N, Martin D. Machine learning and network analyses reveal disease subtypes of pancreatic cancer and their molecular characteristics. Sci Rep. 2020;10(1):1212. doi: 10.1038/s41598-020-58290-2
30. Jeon J, Nim S, Teyra J, et al. A systematic approach to identify novel cancer drug targets using machine learning, inhibitor design and high-throughput screening. Genome Med. 2014;6(7):1–18. doi: 10.1186/s13073-014-0057-7
31. Tong Z, Zhou Y, Wang J. Identifying potential drug targets in hepatocellular carcinoma based on network analysis and one-class support vector machine. Sci Rep. 2019;9(1):1–9. doi: 10.1038/s41598-019-46540-x
32. Raies A, Tulodziecka E, Stainer J, et al. DrugnomeAI Is an ensemble machine-learning framework for predicting druggability of candidate drug targets. Commun Biol. 2022;5(1):1291. doi: 10.1038/s42003-022-04245-4
33. Nada H, Kim S, Lee K. PT-finder: a multi-modal neural network approach to target identification. Comput Biol Med. 2024;174:108444. doi: 10.1016/j.compbiomed.2024.108444
34. Gfeller D, Grosdidier A, Wirth M, et al. SwissTargetPrediction: a web server for target prediction of bioactive small molecules. Nucleic Acids Res. 2014;42(W1):W32–W38. doi: 10.1093/nar/gku293
35. Mouchlis VD, Afantitis A, Serra A, et al. Advances in de novo drug design: from conventional to machine learning methods. Int J Mol Sci. 2021;22(4):1676. doi: 10.3390/ijms22041676
36. Segler MHS, Kogej T, Tyrchan C, et al. Generating focused molecule libraries for drug discovery with recurrent neural networks. ACS Cent Sci. 2018;4(1):120–131. doi: 10.1021/acscentsci.7b00512.

- • **This study showcases the use of RNNs as a generative model for de novo drug design.**

37. Olivecrona M, Blaschke T, Engkvist O, et al. Molecular De-novo design through deep reinforcement learning. J Cheminformatics. 2017;9(1):1–14. doi: 10.1186/s13321-017-0235-x
38. Popova M, Isayev O, Tropsha A. Deep reinforcement learning for de novo drug design. Sci Adv. 2018;4(7):eaap7885. doi: 10.1126/sciadv.aap7885
39. Bai Q, Liu S, Tian Y, et al. Application advances of deep learning methods for de novo drug design and molecular dynamics simulation. Wiley Interdiscip Rev Comput Mol Sci. 2022;12(3):3. doi: 10.1002/wcms.1581
40. Li Y, Zhang L, Liu Z. Multi-objective de novo drug design with conditional graph generative model. J Cheminformatics. 2018;10 (1):1–24. doi: 10.1186/s13321-018-0287-6
41. Khemchandani Y, O'Hagan S, Samanta S, et al. Computational strategy for generating molecules with desirable properties: a graph convolution and reinforcement learning approach. J Cheminformatics. 2020;12(1):1–17. doi: 10.1186/s13321-020-00454-3
42. You J, Liu B, Ying Z, et al. Graph convolutional policy network for goal-directed molecular graph generation. 32nd Conference on Neural Information Processing Systems (NeurIPS 2018); Montréal, Canada; 2018.p. 31.
43. Putin E, Asadulaev A, Ivanenkov Y, et al. Reinforced adversarial neural computer for de novo molecular design. J Chem Inf Model. 2018;58(6):1194–1204. doi: 10.1021/acs.jcim.7b00690
44. Guimaraes GL, Sanchez-Lengeling B, Outeiral C, et al. Objective-reinforced generative adversarial networks (organ) for sequence generation models. ArXiv Prepr. ArXiv170510843. 2017.
45. Sanchez-Lengeling B, Outeiral C, Guimaraes GL, et al. Optimizing distributions over molecular space. an objective-reinforced

generative adversarial network for inverse-design chemistry (ORGANIC). ChemRxiv. 2017. doi: 10.26434/chemrxiv.5309668.v3

46. De Cao N, Kipf T. MolGAN: an implicit generative model for small molecular graphs. ArXiv Prepr. ArXiv180511973. 2018.
47. Putin E, Asadulaev A, Vanhaelen Q, et al. Adversarial threshold neural computer for molecular de novo design. Mol Pharm. 2018;15(10):4386–4397. doi: 10.1021/acs.molpharmaceut.7b01137
48. Gómez-Bombarelli R, Wei JN, Duvenaud D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268–276. doi: 10.1021/acscentsci.7b00572
49. Lim J, Ryu S, Kim JW, et al. Molecular generative model based on conditional variational autoencoder for de novo molecular design. J Cheminformatics. 2018;10(1):1–9. doi: 10.1186/s13321-018-0286-7
50. Born J, Manica M, Oskooei A, et al. PaccMannRL: de novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning. iScience. 2021;24(4):102269. doi: 10.1016/j.isci.2021.102269
51. Bagal V, Aggarwal R, Vinod PK, et al. MolGPT: molecular generation using a transformer-decoder model. J Chem Inf Model. 2022;62(9):2064–2076. doi: 10.1021/acs.jcim.1c00600
52. Kadurin A, Aliper A, Kazennov A, et al. The cornucopia of meaningful leads: applying deep adversarial autoencoders for new molecule development in oncology. Oncotarget. 2017;8(7):10883–10890. doi: 10.18632/oncotarget.14073
53. Kadurin A, Nikolenko S, Khrabrov K, et al. An advanced generative adversarial autoencoder model for de novo generation of new molecules with desired molecular properties in silico. Mol Pharm. 2017;14(9):3098–3104. doi: 10.1021/acs.molpharmaceut.7b00346
54. Prykhodko O, Johansson SV, Kotsias P-C, et al. A de novo molecular generation method using latent vector based generative adversarial network. J Cheminformatics. 2019;11(1):1–13. doi: 10.1186/s13321-019-0397-9
55. Li Y, Liu Y, Wu J, et al. Discovery of potent, selective, and orally bioavailable small-molecule inhibitors of CDK8 for the treatment of cancer. J Med Chem. 2023;66(8):5439–5452. doi: 10.1021/acs.jmedchem.2c01718
56. Lyne PD. Structure-based virtual screening: an overview. Drug Discov Today. 2002;7(20):1047–1055. doi: 10.1016/S1359-6446(02)02483-2
57. Oliveira TAD, Silva MPD, Maia EHB, et al. Virtual screening algorithms in drug discovery: a review focused on machine and deep learning methods. Drugs Drug Candidates. 2023;2(2):311–334. doi: 10.3390/ddc2020017
58. Ain QU, Aleksandrova A, Roessler FD, et al. Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. Wiley Interdiscip Rev Comput Mol Sci. 2015;5(6):405–424. doi: 10.1002/wcms.1225
59. Dara S, Dhamercherla S, Jadav SS, et al. Machine learning in drug discovery: a review. Artif Intell Rev. 2022;55(3):1947–1999. doi: 10.1007/s10462-021-10058-4
60. Wójcikowski M, Ballester PJ, Siedlecki P. Performance of machine-learning scoring functions in structure-based virtual screening. Sci Rep. 2017;7(1):1–10. doi: 10.1038/srep46710
61. Li G-B, Yang L-L, Wang W-J, et al. ID-score: a new empirical scoring function based on a comprehensive set of descriptors related to protein–ligand interactions. J Chem Inf Model. 2013;53(3):592–600. doi: 10.1021/ci300493w
62. Durrant JD, McCammon JA. NNScore 2.0: A neural-network receptor–ligand scoring function. J Chem Inf Model. 2011;51(11):2897–2903. doi: 10.1021/ci2003889
63. Wijewardhane PR, Jethava KP, Fine JA, et al. Combined molecular graph neural network and structural docking selects potent programmable cell death protein 1/programmable death-ligand 1 (PD-1/PD-L1 *Small Molecule Inhibitors*; preprint; Chemistry, 2020. doi: 10.26434/chemrxiv.12083907.v1
64. Zhan W, Li D, Che J, et al. Integrating docking scores, interaction profiles and molecular descriptors to improve the accuracy of molecular docking: toward the discovery of novel Akt1 inhibitors. Eur J Med Chem. 2014;75:11–20. doi: 10.1016/j.ejmech.2014.01.019
65. Durrant JD, Carlson KE, Martin TA, et al. Neural-network scoring functions identify structurally novel estrogen-receptor ligands. J Chem Inf Model. 2015;55(9):1953–1961. doi: 10.1021/acs.jcim.5b00241
66. Wang S, Xu F, Li Y, et al. KG4SL: knowledge graph neural network for synthetic lethality prediction in human cancers. Bioinformatics. 2021;37(Supplement_1):i418–i425. doi: 10.1093/bioinformatics/btab271
67. Cerchia C, Lavecchia A. New avenues in artificial-intelligence-assisted drug discovery. Drug Discov Today. 2023;28(4):103516. doi: 10.1016/j.drudis.2023.103516
68. Xie Q-Q, Zhong L, Pan Y-L, et al. Combined SVM-based and docking-based virtual screening for retrieving novel inhibitors of c-met. Eur J Med Chem. 2011;46(9):3675–3680. doi: 10.1016/j.ejmech.2011.05.031
69. Valarmathi T, Premkumar R, Meera MR, et al. Quantum chemical and molecular docking studies on 1-amino-5-chloroanthraquinone: a targeted drug therapy for thyroid cancer. Spectrochim Acta A Mol Biomol Spectrosc. 2021;255:119659. doi: 10.1016/j.saa.2021.119659
70. Adon T, Shanmugarajan D, Ather H, et al. Virtual screening for identification of dual inhibitors against CDK4/6 and aromatase enzyme. Molecules. 2023;28(6):2490. doi: 10.3390/molecules28062490
71. Yu M, Gu Q, Xu J. Discovering new PI3Kα inhibitors with a strategy of combining ligand-based and structure-based virtual screening. J Comput Aided Mol Des. 2018;32(2):347–361. doi: 10.1007/s10822-017-0092-8
72. Fan C, Huang Y. Identification of novel potential scaffold for class I HDACs inhibition: an in-silico protocol based on virtual screening, molecular dynamics, mathematical analysis and machine learning. Biochem Biophys Res Commun. 2017;491(3):800–806. doi: 10.1016/j.bbrc.2017.07.051
73. Ren M, Li D, Liu G, et al. Discovery novel VEGFA inhibitors through structure-based virtual screening and verify the ability to inhibit the proliferation, invasion and migration of gastric cancer. J Saudi Chem Soc. 2023;2023(4):101674. doi: 10.1016/j.jscs.2023.101674
74. Johnson MA, Maggiora GM. Concepts and Applications in Molecular Similarity. (NY): Wiley; 1990.
75. Valentini E, D'Aguanno S, Di Martile M, et al. Targeting the anti-apoptotic bcl-2 family proteins: machine learning virtual screening and biological evaluation of new small molecules. Theranostics. 2022;12(5):2427–2444. doi: 10.7150/thno.64233
76. Di Stefano M, Galati S, Ortore G, et al. Machine learning-based virtual screening for the identification of CDK5 inhibitors. Int J Mol Sci. 2022;23(18):10653. doi: 10.3390/ijms231810653
77. Zhang H, Huang J, Chen R, et al. Ligand- and structure-based identification of novel CDK9 inhibitors for the potential treatment of leukemia. Bioorg Med Chem. 2022;72:116994. doi: 10.1016/j.bmc.2022.116994
78. Bahi M, Batouche M. Deep learning for ligand-based virtual screening in drug discovery. In: 2018 3rd International Conference on Pattern Analysis and Intelligent Systems (PAIS). IEEE: Tebessa; 2018. p. 1–5. doi: 10.1109/PAIS.2018.8598488
79. Xu Y, Chen P, Lin X, et al. Discovery of CDK4 inhibitors by convolutional neural networks. Future Med Chem. 2019;11(3):165–177. doi: 10.4155/fmc-2018-0478
80. Lu M, Yin J, Zhu Q, et al. Artificial intelligence in pharmaceutical sciences. Eng. 2023:S2095809923001649. doi: 10.1016/j.eng.2023.01.014
81. Kimber TB, Chen Y, Volkamer A. Deep learning in virtual screening: recent applications and developments. Int J Mol Sci. 2021;22(9):4435. doi: 10.3390/ijms22094435
82. Pandiyan S, Wang L. A comprehensive review on recent approaches for cancer drug discovery associated with artificial intelligence. Comput Biol Med. 2022;150:106140. doi: 10.1016/j.compbiomed.2022.106140
83. Paul D, Sanap G, Shenoy S, et al. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021;26(1):80–93. doi: 10.1016/j.drudis.2020.10.010

84. Hamdi A, Colas P. Yeast two-hybrid methods and their applications in drug discovery. Trends Pharmacol Sci. 2012;33(2):109–118. doi: 10.1016/j.tips.2011.10.008
85. Omidfar K, Daneshpour M. Advances in phage display technology for drug discovery. Expert opinion on drug discovery. 2015;10 (6):651–669.
86. Wang F, Liu D, Wang H, et al. Computational screening for active compounds targeting Protein sequences: methodology and experimental validation. J Chem Inf Model. 2011;51(11):2821–2828. doi: 10.1021/ci200264h
87. Yu H, Chen J, Xu X, et al. A systematic prediction of multiple drug-target interactions from chemical, genomic, and pharmacological data. PLoS One. 2012;7(5):e37608. doi: 10.1371/journal.pone.0037608
88. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. Bioinformatics. 2018;34(17):i821–i829. doi: 10.1093/bioinformatics/bty593
89. Wen M, Zhang Z, Niu S, et al. Deep-learning-based drug–target interaction prediction. J Proteome Res. 2017;16(4):1401–1409. doi: 10.1021/acs.jproteome.6b00618
90. Lee I, Keum J, Nam H. DeepConv-DTI: prediction of drug-target interactions via deep learning with convolution on protein sequences. PLoS Comput Biol. 2019;15(6):e1007129. doi: 10.1371/journal.pcbi.1007129
91. Gao KY, Fokoue A, Luo H, et al. Interpretable drug target prediction using deep neural representation. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence; International Joint Conferences on Artificial Intelligence Organization; Stockholm, Sweden; 2018. p. 3371–3377. doi: 10.24963/ijcai.2018/468
92. Rifaioglu AS, Atas H, Martin MJ, et al. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. Brief Bioinform. 2019;20 (5):1878–1912. doi: 10.1093/bib/bby061
93. Wang C, Liu J, Luo F, et al. Pairwise input neural network for target-ligand interaction prediction. In: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Belfast, UK. IEEE; 2014. p. 67–70. doi: 10.1109/BIBM.2014.6999129
94. Hinnerichs T, Hoehndorf R, Wren J. DTI-voodoo: machine learning over interaction networks and ontology-based background knowledge predicts drug–target interactions. Bioinformatics. 2021;37(24):4835–4843. doi: 10.1093/bioinformatics/btab548

•• **This study presents a transformer-based VS workflow for discovering CDK12 inhibitors in cancers therapeutics.**

95. Wen T, Wang J, Lu R, et al. Validation, and evaluation of a deep learning model to screen cyclin-dependent kinase 12 inhibitors in cancers. Eur J Med Chem. 2023;2023:115199. doi: 10.1016/j.ejmech.2023.115199
96. Nguyen T, Le H, Quinn TP, et al. GraphDTA: predicting drug–target binding affinity with graph neural networks. Bioinformatics. 2021;37(8):1140–1147. doi: 10.1093/bioinformatics/btaa921
97. Nguyen TM, Nguyen T, Le TM, et al. Gefa: early fusion approach in drug-target affinity prediction. IEEE/ACM Trans Comput Biol Bioinform. 2021;19(2):718–728. doi: 10.1109/TCBB.2021.3094217
98. Yang Z, Zhong W, Zhao L, et al. MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction. Chem Sci. 2022;13(3):816–833. doi: 10.1039/D1SC05180F
99. Zang Q, Mansouri K, Williams AJ, et al. In silico prediction of physicochemical properties of environmental chemicals using molecular fingerprints and machine learning. J Chem Inf Model. 2017;57(1):36–49. doi: 10.1021/acs.jcim.6b00625
100. Panapitiya G, Girard M, Hollas A, et al. Evaluation of deep learning architectures for aqueous solubility prediction. ACS Omega. 2022;7(18):15695–15710. doi: 10.1021/acsomega.2c00642
101. Kamlet MJ, Doherty RM, Abboud J-LM, et al. Linear solvation energy relationships: 36. molecular properties governing solubilities of organic nonelectrolytes in water. J Pharm Sci. 1986;75 (4):338–349. doi: 10.1002/jps.2600750405
102. Lusci A, Pollastri G, Baldi P. Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules. J Chem Inf Model. 2015:28.
103. Francoeur PG, Koes DR. SolTranNet–A machine learning tool for fast aqueous solubility prediction. J Chem Inf Model. 2021;61 (6):2530–2536. doi: 10.1021/acs.jcim.1c00331
104. Suenderhauf C, Hammann F, Maunz A, et al. Combinatorial QSAR modeling of human intestinal absorption. Mol Pharm. 2011;8 (1):213–224. doi: 10.1021/mp100279d
105. Dixon SL, Duan J, Smith E, et al. AutoQSAR: an automated machine learning tool for best-practice quantitative structure–activity relationship modeling. Future Med Chem. 2016;8(15):1825–1839. doi: 10.4155/fmc-2016-0093
106. Wang N-N, Dong J, Deng Y-H, et al. ADME properties evaluation in drug discovery: prediction of caco-2 cell permeability using a combination of NSGA-II and boosting. J Chem Inf Model. 2016;56(4):763–773. doi: 10.1021/acs.jcim.5b00642
107. Zhou T, Jhamb S, Liang X, et al. Prediction of acid dissociation constants of organic compounds using group contribution methods. Chem Eng Sci. 2018;183:95–105. doi: 10.1016/j.ces.2018.03.005
108. Cumming JG, Davis AM, Muresan S, et al. Chemical predictive modelling to improve compound quality. Nat Rev Drug Discov. 2013;12(12):948–962. doi: 10.1038/nrd4128
109. Ramana PV, Krishna YR, Mouli KC. Experimental FT-IR and UV–Vis spectroscopic studies and molecular docking analysis of anti-cancer drugs exemestane and pazopanib. J Mol Struct. 2022;2022:133051. doi: 10.1016/j.molstruc.2022.133051
110. Moroy G, Martiny VY, Vayer P, et al. Toward in silico structure-based ADMET prediction in drug discovery. Drug Discov Today. 2012;17 (1–2):44–55. doi: 10.1016/j.drudis.2011.10.023
111. Bisson WH, Cheltsov AV, Bruey-Sedano N, et al. Discovery of antiandrogen activity of nonsteroidal scaffolds of marketed drugs. Proc Natl Acad Sci. 2007;104(29):11927–11932. doi: 10.1073/pnas.0609752104
112. Jimenez-Carretero D, Abrishami V, Fernández-de-Manuel L, et al. Tox_(R)CNN: deep learning-based nuclei profiling tool for drug toxicity screening. PLoS Comput Biol. 2018;14(11):e1006238. doi: 10.1371/journal.pcbi.1006238
113. Daina A, Michielin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci Rep. 2017;7(1):42717. doi: 10.1038/srep42717
114. Fu L, Shi S, Yi J, et al. ADMETlab 3.0: an updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support. Nucleic Acids Res. 2024:gkae236. doi: 10.1093/nar/gkae236

•• **This article presents ADMETlab 3.0, an updated online platform providing comprehensive evaluation of ADMET-related parameters in drug discovery.**

115. Yi J, Shi S, Fu L, et al. OptADMET: a web-based tool for substructure modifications to improve ADMET properties of lead compounds. Nat Protoc. 2024;19(4):1105–1121. doi: 10.1038/s41596-023-00942-4
116. Li B, Dai C, Wang L, et al. A novel drug repurposing approach for non-small cell lung cancer using deep learning. PLoS One. 2020;15(6):e0233112. doi: 10.1371/journal.pone.0233112
117. Zeng X, Zhu S, Lu W, et al. Target identification among known drugs by deep learning from heterogeneous networks. Chem Sci. 2020;11(7):1775–1797. doi: 10.1039/C9SC04336E
118. Cheng F, Zhao J, Fooksa M, et al. A network-based drug repositioning infrastructure for precision cancer medicine through targeting significantly mutated genes in the human cancer genomes. J Am Med Inform Assoc. 2016;23(4):681–691. doi: 10.1093/jamia/ocw007
119. Napolitano F, Zhao Y, Moreira VM, et al. Drug repositioning: a machine-learning approach through data integration. J Cheminformatics. 2013;5(1):1–9. doi: 10.1186/1758-2946-5-30
120. Cui C, Ding X, Wang D, et al. Drug repurposing against breast cancer by integrating drug-exposure expression profiles and

drug–drug links based on graph neural network. Bioinformatics. 2021;37(18):2930–2937. doi: 10.1093/bioinformatics/btab191

121. Iorio F, Bosotti R, Scacheri E, et al. Discovery of drug mode of action and drug repositioning from transcriptional responses. Proc Natl Acad Sci. 2010;107(33):14621–14626. doi: 10.1073/pnas.1000138107

122. Persidis A. The benefits of drug repositioning. Drug Discovery World. 2011;12:9–12.

123. Bhinder B, Gilvary C, Madhukar NS, et al. Artificial intelligence in cancer research and precision medicine. Cancer Discov. 2021;11(4):900–915. doi: 10.1158/2159-8290.CD-21-0090

124. Patel V, Shah M. Artificial intelligence and machine learning in drug discovery and development. Intell Med. 2022;2(3):134–140. doi: 10.1016/j.imed.2021.10.001

125. Margulis E, Dagan-Wiener A, Ives RS, et al. Intense bitterness of molecules: machine learning for expediting drug discovery. Comput Struct Biotechnol J. 2021;19:568–576. doi: 10.1016/j.csbj.2020.12.030

126. Ali M, Aittokallio T. Machine learning and feature selection for drug response prediction in precision oncology applications. Biophys Rev. 2019;11(1):31–39. doi: 10.1007/s12551-018-0446-z

127. Dawood M, Vu QD, Young LS, et al. Cancer drug sensitivity prediction from routine histology images. NPJ Precis Oncol. 2024;8(1):5. doi: 10.1038/s41698-023-00491-9.
    - **This proof-of-concept study demonstrates the use of DL to predict cancer drug sensitivity from histological patterns in whole slide images of breast cancer sections.**

128. Huang C, Mezencev R, McDonald JF, et al. Open source machine-learning algorithms for the prediction of optimal cancer drug therapies. PLoS One. 2017;12(10):e0186906. doi: 10.1371/journal.pone.0186906

129. Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463–477. doi: 10.1038/s41573-019-0024-5

130. Chen W, Liu X, Zhang S, et al. Artificial intelligence for drug discovery: resources, methods, and applications. Mol Ther Nucleic Acids. 2023;31:691–702. doi: 10.1016/j.omtn.2023.02.019

131. Bhat AA, Nisar S, Mukherjee S, et al. Integration of CRISPR/Cas9 with artificial intelligence for improved cancer therapeutics. J Transl Med. 2022;20(1):534. doi: 10.1186/s12967-022-03765-1