npj | precision oncology

Review

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-026-01310-7

AI accelerate the identification of druggable targets by 3D structures of proteins and compounds

Da Li[1,2], Sanbao Shi[1,2], Zhiyu Yu[1], Peng Xu[1] ✉ & Cheng Zhang[1] ✉

Artificial intelligence (AI) is being used in oncological drug development to address the high costs, low success rates, and long timelines that characterize traditional drug development pipelines. The use of machine learning (ML) and deep learning (DL) models in computer-aided drug design is constantly growing owing to their capacity to analyze large, heterogeneous datasets, their ability to capture nonlinear biological trends, and their integration of various molecular and clinical characteristics. AI applications accelerate target discovery by predicting protein structures, ranking disease-relevant genes, and assessing target drugability. AI can be used to conduct rapid searches of multiplexed chemical libraries, predict drug-target interactions, and optimize the pharmacological and physicochemical properties of drugs in virtual screening. Advanced neural network designs also aid in de novo drug design, which involves developing new molecular structures with therapeutic properties of interest. This review outlines how AI has been used for target identification, virtual screening, de novo molecular design, and, specifically, in cancer applications. It further discusses the major issues in AI-based drug development, such as data quality, model interpretation, computational constraints, and ethical and regulatory considerations, which remain essential obstacles to broader clinical translation.

Drug discovery is the process of identifying new medicinal agents through testing, computation, and experimentation[1]. It begins with the discovery of biomolecules and pathways, which are subject to change to generate therapeutic benefits. After selecting four to six targets, thousands of compounds are screened to identify first-generation hits with decent affinity using high-throughput or virtual screening techniques[2]. Optimization is then performed to enhance the pharmacokinetic, therapeutic, and safety characteristics of the lead compounds with the best potency and selectivity. Such candidates are refined and subjected to preclinical and clinical testing, with the aim of regulatory approval. The entire process may take 12–15 years and cost USD 2.8 billion[3].

The development of improved techniques has been enabled by technological advances such as computer-aided drug design (CADD), automation, and nanofluidics, which have enhanced their accuracy and efficiency. Nevertheless, an incomplete understanding of disease pathologic processes, constraints of preclinical models, patient heterogeneity, and the absence of predictive biomarkers continue to limit studies[4,5]. All of these issues are especially acute in oncology, where the number of druggable targets is minimal, tumours are heterogeneous, and chemoresistance is common[6].

From conventional approaches to CADD

Computer-Aided Drug Design in the 1990s changed this when it began using computational tools to construct in silico models of drug-target interactions, which were then subjected to laboratory testing[7]. Contrary to the old system of trial and error, CADD introduces a more rational approach, which speeds development, reduces experimental requirements, and makes the drug more cost-effective[8]. The incorporation of big data and multi-omics has also led CADD to become a multidisciplinary field encompassing systems biology, mathematical modelling, computer simulations, bioinformatics, and pattern recognition. These techniques have been helpful in identifying targets, modelling drug-target interactions, performing virtual screening, and optimizing leads[9]. However, the nonlinear, high-dimensional relationships in contemporary biomedical data are often poorly suited to conventional CADD, which relies on predefined algorithms[10].

[1]Department of General Surgery, General Hospital of Northern Theater Command, Shenyang, Liaoning Province, China. [2]These authors contributed equally: Da Li, Sanbao Shi. ✉e-mail: pppengxu@163.com; zhangc1109@163.com

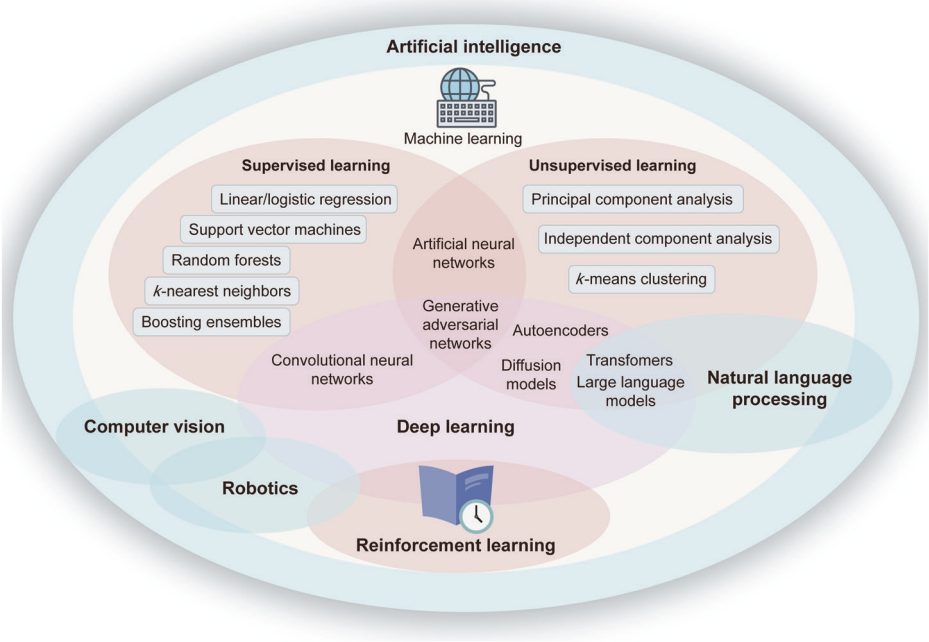

Fig. 1 | Hierarchical overview of AI, ML, and DL. Emphasizing the principal architectures of supervised, unsupervised, and reinforcement learning, as well as neural networks, employed in computational drug discovery. Adapted from reference[16].

Emergence of AI in drug discovery

The solution to these issues with the help of AI is robust. AI systems mimic human intelligence to perform tasks in learning, reasoning, and decision-making[11]. ML is an essential element of AI, enabling models to learn from datasets[12]. Supervised learning uses labelled datasets to generate predictive models for classification and regression, whereas unsupervised learning aims to identify hidden patterns in unlabelled data. Reinforcement learning (RL) improves decision-making by providing trial-and-error information. One subdivision of ML, DL, involves applying multilayer neural networks to extract complex patterns from extensive, unstructured data[13].

Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNNs), generative adversarial networks (GANs), recurrent neural networks (RNNs), variational autoencoders (VAEs) are DL frameworks that have demonstrated high efficacy in biomedical tasks such as medical imaging, medical diagnostics, and natural language processing[14]. Their ability to learn continuously and correct their errors makes them especially successful at analyzing the diverse, complex datasets that characterize drug discovery. Figure 1. Hierarchical hierarchy between AI, ML, and DL, for example, the major algorithmic classes applied in drug discovery[15,16]. The basis of ML is traditional supervised and unsupervised learning, combined with state-of-the-art DL architectures, including CNNs, GANs, diffusion models, transformers, and large language models, that model complex, high-dimensional patterns. The figure also illustrates links to natural language processing, computer vision, robotics, and reinforcement learning, placing these techniques within the broader AI ecosystem of drug discovery[16].

AI as an extension of CADD

ML and DL approaches are currently used to supplement and complement traditional CADD[17]. These techniques enhance predictive accuracy by training on large volumes of data, thereby eliminating the need for pre-existing guidelines[18]. AI has been implemented at almost all stages of drug development, including target discovery, virtual screening, lead optimization, preclinical validation, and some aspects of clinical development[19,20]. AI applications can predict protein structures, screen millions of compounds, predict drug-target interactions, and optimize molecular characteristics[21,22]. DL architectures, including VAEs, GANs, and RL systems, can produce new molecular structures with specific, desirable therapeutic properties in the absence of template information for de novo drug design[23,24].

According to recent reports, the role of AI is growing, and over 150 AI-designed small-molecule programs are in preclinical development, with several already reaching clinical trials[25–27]. The success rates in the early clinical stages of AI-inspired biotechnology firms have exceeded historical standards, and development schedules and research and development (R&D) expenses are dramatically lower[28].

Transitional summary and scope of this review

Although AI has proven very useful for identifying targets and enhancing lead optimization, most AI-based drug discovery companies now focus on validated targets[29,30]. Conversely, large pharmaceutical firms are also making larger investments in new drug targets. However, AI has demonstrated its ability to produce hitherto unknown chemical compounds with astonishing speed and accuracy[31].

This review discusses the application of AI in early oncology drug discovery, focusing on target discovery, virtual screening, and de novo design. The analysis focused on cancer-specific case studies. We also discuss several significant constraints, including data quality, model interpretability, computational constraints, and ethical considerations, that still hinder wider application. The conceptual differences between the workflows of conventional drug discovery and innovative AI-based methods are depicted in Fig. 2[32].

This work makes several distinctive contributions that could advance the field of AI deployment in oncology drug discovery[33]. First, unlike the rest of the reviews that discuss AI tools in isolation, we introduce a logical conceptual framework, biomapping AI methods along the entire early-stage oncology pipeline, comprising target identification, virtual screening, and de novo molecular design, to demonstrate how these methods rely on each other, not to mention that these approaches have a standard methodology. Second, we plotted a new comparative taxonomy of AI algorithms, particularly for oncological applications, that integrates structural prediction algorithms, graph models, multimodal learning systems, and generative structures in a sequential grid of their data requirements, computational requirements, and disease-specific utility. Third, we focus on cancer-specific primary research rather than generic applications of AI by presenting detailed case studies (e.g., AlphaFold-enabled target discovery, GNN-based screening, and multi-omics-conditioned generative models) to demonstrate how these applications are being applied in actual experimental and preclinical oncology settings. The limitations of embedding oncology-specific interactions, translational bottlenecks, and reproducibility issues are critically evaluated, and a future perspective is provided on how multimodal foundation models, XAI frameworks, and closed-loop discovery systems

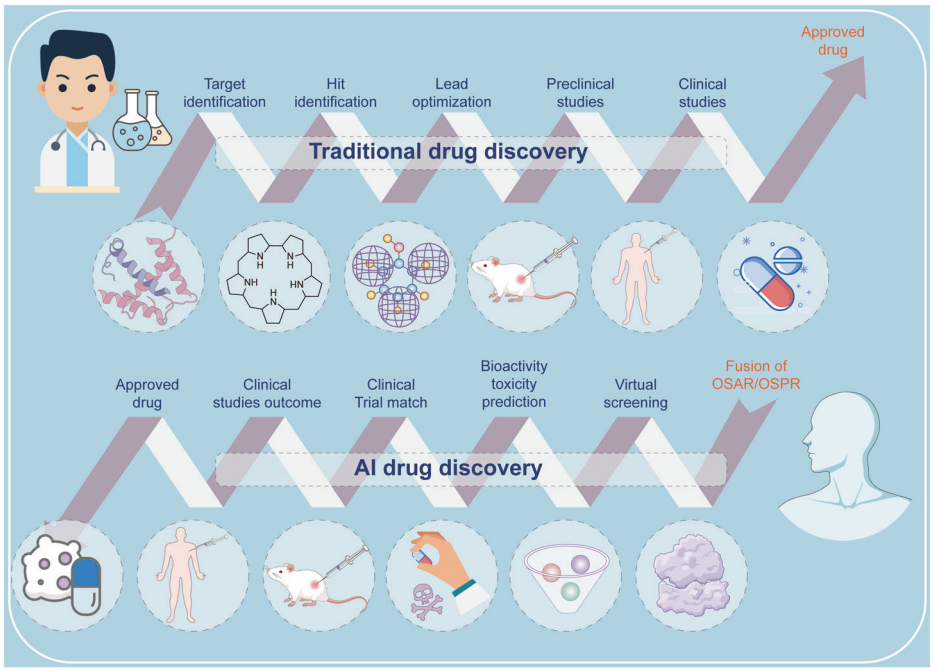

Fig. 2 | Comparative analysis of traditional and AI-driven drug-discovery workflows. The conventional approach to drug discovery follows a sequential progression through experimental stages, beginning with target identification and culminating in clinical evaluation. In contrast, AI-driven methodologies integrate predictive modelling, QSAR/QSPR analysis, toxicity prediction, and virtual screening, thereby expediting early decision-making and reducing experimental workload. Adapted from reference[32].

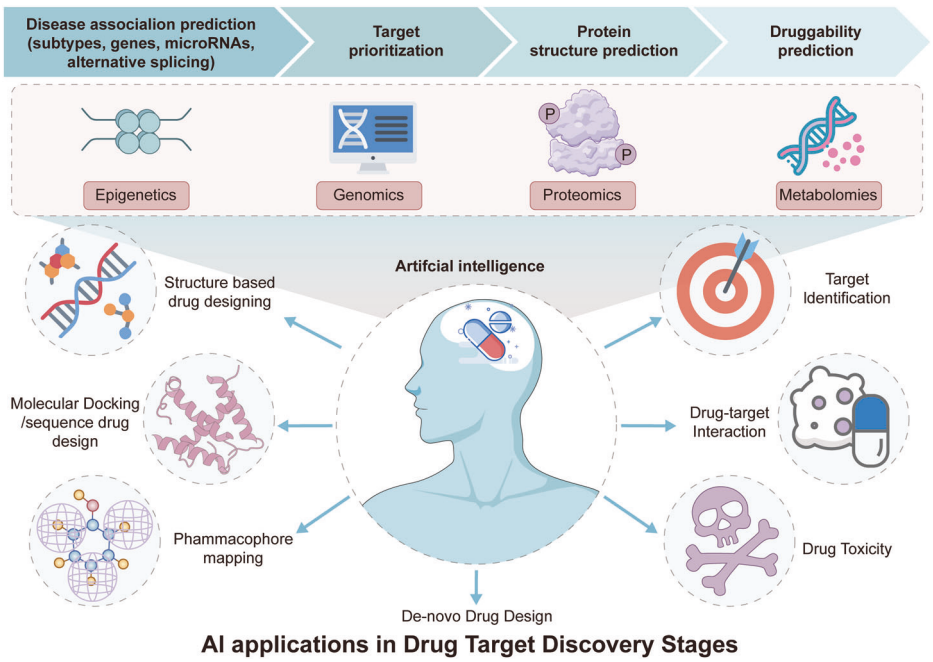

Fig. 3 | Artificial intelligence applications in drug target discovery stages. The diagram illustrates the integration of AI across key phases of target discovery, including disease association prediction (subtypes, genes, microRNAs, alternative splicing), target prioritization, protein structure prediction, and druggability prediction. Multi-omics data sources (epigenetics, genomics, proteomics, and metabolomics) feed into AI-driven analyses for target identification, drug-target interaction prediction, and toxicity assessment. AI facilitates structure-based drug design through molecular docking, pharmacophore mapping, and de novo drug design. The central role of artificial intelligence in integrating these diverse computational and experimental approaches is highlighted.

can redefine future cancer treatment. These factors make this review a valuable synthesis of oncology-specific knowledge, both methodological and translational.

AI-driven target identification

Target identification is a critical first step in the drug discovery process and entails the identification of biomolecules or pathways whose regulation can be therapeutically beneficial. Drug targets are currently discovered using experimental approaches, including chemical and genetic screening, affinity-based screening, and multi-omics analysis, as well as computational techniques such as molecular docking and AI[9]. As cancer is a highly genetic, molecular, and immunological complex disease, multi-source data are highly likely to succeed in identifying therapeutic targets[34,35]. Figure 3. Summary of AI applications in drug target discovery phases, such as disease association prediction, target prioritization, structural prediction, and druggability evaluation[36].

Because of their ability to process large volumes of multi-omics data, identify complex oncogenic networks, and detect novel targets that conventional approaches may miss, AI models are becoming a potent tool for in silico cancer target identification[37,38]. ML and DL algorithms can analyze genomic and transcriptomic data to identify recurrent mutations and expression patterns in cancer, and integrate them with proteomic and epigenomic data to place dysregulated proteins and pathways at the top of the priority list[39]. The integration of clinical data will also enable the correlation of molecular characteristics with treatment response, contributing to the formulation of individualized therapy. Scientific literature, clinical

trial information, and patent data can be mined to discover disease-gene-drug associations and targets using AI-based natural language processing systems, such as LLMs (e.g., BioGPT, ChatGPT, ChatPanda)[26,40]. Large-scale AI and data modalities will be applied to target discovery in oncology, as shown in Tables 1 and 2.

An example of this approach is PandaOmics, an AI-based multi-omics and text-mining platform created by Insilico Medicine[37,41]. It integrates genomic, transcriptomic, proteomic, and methylomic data over literature-based knowledge graphs to determine and rank potential therapeutic targets[42]. GWAS, differential expression, and mutational analyses (Fig. 4) were used to create omics-based scores, which were then used to rank the diseases and select the most relevant one, based on text-based scores reflecting the disease's relevance in the published literature[43]. Schematic workflow of the PandaOmics and Chemistry42 platforms to integrate multi-omics, identify targets, rank them based on prior knowledge graphs, and design small-molecules generatively[41,44]. PandaOmics has been shown to be effective across most cancer types, including CNGA3, GLUD1, and SIRT1 targets in glioblastoma; CDK20 in hepatocellular carcinoma; and CDK12 in gastric, colorectal, and triple-negative breast cancers[41,45,46].

Druggability, which is the likelihood of a target being altered by a small molecule, is vital for drug development[9,47]. Many ML and DL models have been trained to forecast protein druggability using sequence-based, structural, topological, and physicochemical characteristics[48]. Among these are hybrid convolutional neural network-recurrent neural network (CNN-RNN) classifiers trained on DrugBank and Swiss-Prot proteins and DrugnomeAI[49,50]. This semi-supervised system combines protein-protein interaction data, pathways, and protein features to provide drugs with scores of how easy they are to target[49]. Further, network-based methods, including random forests (RFs) trained on protein similarity networks, have been used to predict potential cancer drug targets with high predictive accuracy[51,52].

AI plays a key role in determining the effects of point mutations on protein function, which is a vital component of oncology[53]. Other tools, such as AUTO-MUTE 2.0, combine ML with structural modelling to infer the functional consequences of amino acid mutations[54,55]. AlloDriver builds on this approach by mapping mutations to orthostatic and allosteric sites that are actionable cancer drivers[56,57]. This strategy identified clinically meaningful cases, including an allosteric mutation that is clinically significant, L1143F in PTPRK, which is linked to head and neck cancer[58].

Predicting synthetic lethal (SL) gene pairs has a high potential of being made by AI, as these pairs are good targets of therapeutic intervention[58,59]. Although experimental techniques such as CRISPR/Cas9 and RNA interference (RNAi) screens can be used to determine SL interactions, their use is constrained by high costs and off-target effects[60,61]. AI-based models, including KG4SL, leverage graph neural networks and knowledge graphs from databases such as SynLethDB to improve the accuracy and scalability of SL interaction predictions[62].

AlphaFold2, AlphaFold3, and ESMFold are transformer-based structural prediction models used to predict protein targets by accurately modelling protein structure, analyzing binding pockets, and analyzing structural changes induced by mutations[27,63]. Such developments have expanded the prioritization of AI frameworks for druggable cancer targets in preclinical and translational research.

In oncology, AI has significantly accelerated target identification and validation by leveraging multi-omics data, structural prediction, computational mutagenesis, and literature-based knowledge extraction[9,64].

Virtual screening and lead optimization using AI

Systemic screening of pharmacological compounds with therapeutic potential is an essential step in drug discovery[65]. High-throughput screening (HTS) uses robotics, assays, and data-processing systems to screen hundreds of thousands of compounds against biological targets[66,67]. Although HTS is the main approach for identifying hits, it has certain limitations, including high cost, long development time, and a low hit rate of 1%[68]. Virtual screening helps overcome many of these limitations by applying computational methods to screen compounds in large in silico libraries and

Table 1 | Summary of key artificial intelligence platforms for target identification

Platform	Algorithm type/core model	Primary data types used	Main function in drug discovery	Reference
PandaOmics (Insilico Medicine)	DL ensemble; GAN-based multi-omics integration; knowledge-graph mining	Transcriptomics, proteomics, methylomics, Genome-wide association studies (GWAS), somatic/germline mutations, clinical text	Target identification and prioritization; disease mechanism discovery	44,172
DrugnomeAI (AstraZeneca)	Semi-supervised ML; graph-based learning on protein networks	Protein–protein interactions, pathway data, gene-level annotations (Pharos, TCRD, Triage datasets)	Druggability prediction across the human exome; identification of novel oncology targets	173
AlloDriver	Random Forest + neural network models for allosteric site mapping	Protein 3D structures, mutation profiles, orthosteric/allosteric site features	Identification of oncogenic allosteric driver mutations; prediction of allosteric druggable pockets	56,174
KG4SL	Knowledge graph + graph neural network (GNN) architecture	Integrated heterogeneous biological networks (genes, proteins, drugs, pathways, biological processes)	Prediction of synthetic-lethal gene pairs; discovery of combinatorial cancer targets	175

Table 2 | Algorithms and data types used by AI target identification platforms

Platform	ML/DL algorithms	Key computational approaches	Data modalities integrated	Reference
PandaOmics	GANs, DL classifiers, ensemble ML	Multi-omics scoring; text-mining; knowledge-graph inference	RNA-seq, proteomics, methylomics, GWAS, somatic mutations, clinical trial data, literature	176–179
DrugnomeAI	Semi-supervised ML, graph ML	Protein network embedding; exome-wide druggability scoring	PPI networks, gene–disease associations, protein domains, pathway connectivity	180,181
AlloDriver	Random Forest, deep neural networks	Mutation mapping onto 3D structures; allosteric pocket prediction	Protein 3D structures, cancer somatic mutation databases, allosteric site annotations	182,183
KG4SL	Graph neural networks (GNN), knowledge-graph embedding	Multi-relational graph learning; synthetic-lethality scoring	Gene–gene networks, drug–target networks, pathway maps, biological process ontologies	184

predict drug-target interactions[69,70]. Its capacity is in the millions or even billions of compounds, which is in excess of that of HTS. Chemical libraries are further screened using screening filters to a manageable size that can be experimentally validated, thus saving time and cost, consuming fewer resources, and providing promising leads[71]. In general, virtual screening methods are classified into structure- and ligand-based methods (Fig. 5)[72]. Structure-based screening requires knowledge of the target protein's three-dimensional structure, whereas ligand-based screening identifies compounds with a similar structure or pharmacological properties to known active molecules[73–75].

Conventional virtual screening strategies are susceptible to the computational burden of screening ultra-large chemical libraries[76]. AI models can be beneficial for these processes because they can quickly weed out unsuitable candidates and screen millions of compounds in a limited time[77,78]. DL systems, including CNNs and GNNs, trained on chemical repositories (e.g., PubChem, ChEMBL) can capture nonlinear, including complex, features in molecular structures and biological activity, enabling more accurate predictions of drug-target interactions. The following subsections provide a summary of the AI application in structure- and ligand-based virtual drug screening[79].

AI in structure-based drug screening

Structure-based drug screening uses three-dimensional protein structures to elucidate ligand-target interactions and identify compounds with favourable binding affinities for target proteins[80,81]. Structural determination relies on experimental methodologies such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM)[82]. Molecular docking is commonly used to simulate Ligand binding and estimate binding affinity using scoring functions[83,84]. In contrast, conformational flexibility and MD simulations are frequently used to discover transient or cryptic binding pockets[85]. The results indicate that AI has considerably improved several aspects of structure-based screening, such as structure prediction, scoring functions, and MD simulation efficiency[86,87]. Scoring using AI technologies will be combined with docking workflows, thereby improving affinity predictions and enabling more effective filtering of dormant compounds[88].

Protein structure prediction. Correct knowledge of protein structure is a prerequisite for structure-based drug design (SBDD). The development of AI tools, including AlphaFold, has completely changed the landscape of this field, enabling direct prediction of protein structure from sequence information[89]. AlphaFold2 (AF2) uses a deep neural network called Evoformer, as well as a series of sequence alignments (MSAs) to a database of structural data, to produce protein models with a high level of accuracy[90,91]. AlphaFold3 (AF3) is also much faster and more accurate at modelling and can predict protein-protein and protein-ligand complexes[92].

The use of AF2 as a model has greatly accelerated hit identification, as demonstrated by Ren et al., who used AF2 and the Chemistry42 platform to identify novel CDK20 inhibitors[93,94]. In addition, WSB1 has been inferred from AF2-predicted structures, which have been used in virtual screening campaigns[95]. The computation of allosteric sites has been enhanced by the development of tools such as AlphaFill, which complements AF2 by incorporating cofactors and ligands[96,97]. Other structural prediction systems, including ESMFold, which is not dependent on MSAs, and RoseTTAFold, which combines sequence and spatial information, are also available with slightly different advantages depending on the level of accuracy required and the accessibility of homologous data[98,99]. Overall, these structure prediction tools expand the scope of structure-based drug discovery, enabling more efficient hit identification[99,100].

Molecular docking. Molecular docking is a computational method that estimates the optimal pose and position of ligands when they bind to protein targets[101–103]. The docking process generally involves three major steps: pose generation, scoring, and ranking[104]. New developments in AI

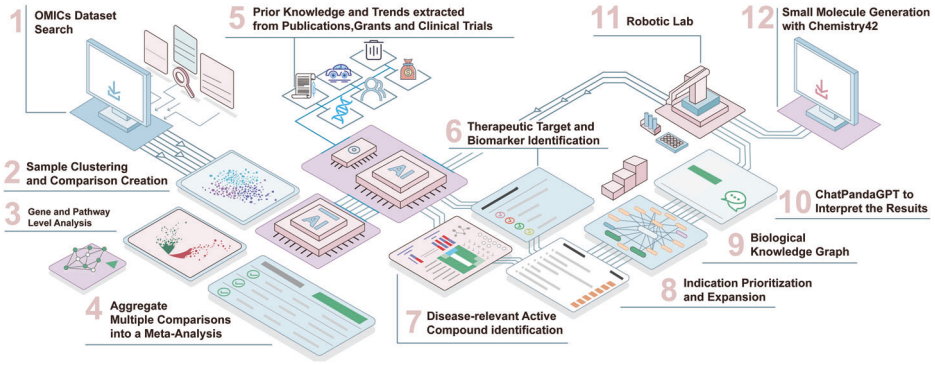

Fig. 4 | PandaOmics–Chemistry pipeline. Comprehensive PandaOmics–Chemistry[172] pipeline illustrates the integration of multi-omics analysis, knowledge graph integration, AI-driven target identification and prioritization, and the generative design of small molecules. Adapted from reference[43].

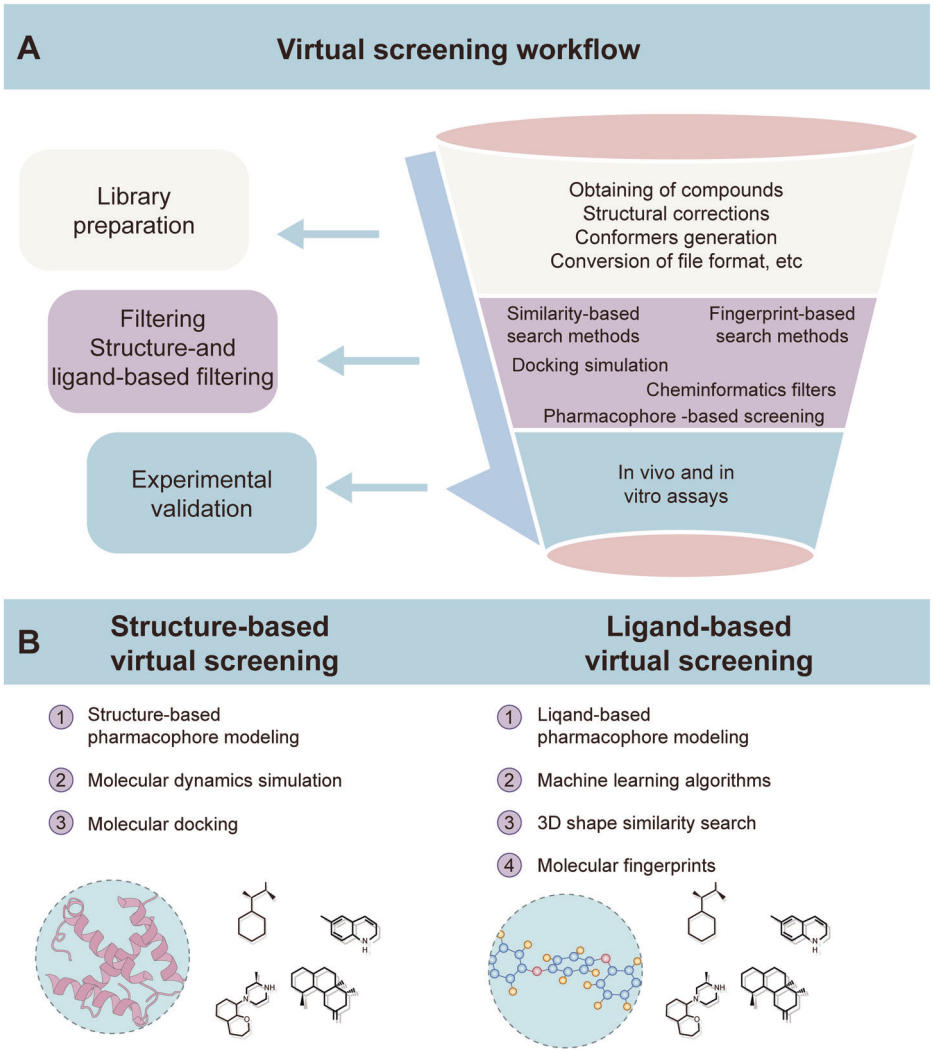

Fig. 5 | Virtual screening funnel illustrating the processes. Virtual screening funnel illustrating the processes of library preparation, structure- and ligand-based filtering, and prioritization of candidate molecules for subsequent experimental validation. Image taken from ref. 72.

have increased the precision of docking using large datasets of protein–ligand complexes[105]. Among them, hybrid support vector machine (SVM)-based methods that utilize molecular descriptors and random forest (RF)-based classifiers that effectively filter out inactive compounds, thereby enhancing overall docking efficiency, should be noted[106].

The recent development of AI-based docking tools is worth mentioning. As a CNN model, AtomNet has been shown to be more effective at predicting the bioactivity of small molecules using atomic-level structural information, achieving significant improvements over traditional docking algorithms in both speed and accuracy[107]. DeepDock, a GNN-based system, can improve docking prediction accuracy by combining features of both ligands and proteins and has been successfully used to identify AXL inhibitors. GNINA was used to restore the binding poses to enhance the ranking of high-quality interactions[108]. Simultaneously, DiffDock uses diffusion models to predict binding sites with high precision and enables "blind docking" that does not presuppose any knowledge of the binding sites[109].

MD simulations. Molecular docking cannot adequately account for the dynamic interactions between ligands and proteins. The limitation of MD simulations is that they model only the time-dependent conformational changes of biomolecules[86,110]. Cryptic pockets can be identified using MD simulations, and these computations can be used to refine docking predictions, which should be verified experimentally[111]. AI can improve MD efficiency by generating accurate force fields, predicting trajectories, and identifying features in large simulation datasets. MDtrajNet and TorchMD: Transformer- and DL-based MD frameworks that are becoming increasingly popular in structure-based drug design are much faster than simulations[112,113].

AI in ligand-based drug screening

The principles of ligand-based screening are that similar molecules should exhibit similar biological activities. This method is advantageous when the target protein's structure is unknown or difficult to determine[114]. Although ligand-based strategies can be used to search large chemical spaces quickly, they can yield false positives because of structural mimicry that is not functionally similar[115]. AI models supplement ligand-based screening by extracting molecular descriptors and predicting biological activity[116].

ML and DL algorithms can detect pharmacophore properties, generate molecular fingerprints, and distinguish active from inactive compounds[117,118]. These models facilitate high-throughput searches for similarities and efficient searches of large databases[119].

Molecular similarity search. Similarity-based screening uses molecular fingerprints to screen based on structural and chemical properties[120]. DL algorithms, specifically GNNs, treat molecules as graphs and produce fingerprints that encode complex topological and spatial data[121]. AI-driven multi-classifier methods within AI-driven workflows have identified EGFR-T790M and NTRK inhibitors, among others[122].

DeepChem, which can be applied to quantitative structure-activity relationship (QSAR) and pharmacophore modelling; PyRMD, which can apply relative molecular descriptor (RMD)-based feature discrimination; and VSFlow, which can be used for similarity searching with the help of RDKit, are automated ligand-based screening platforms[123]. Such tools make screening easier and help find promising ligands faster[124].

QSAR modeling. QSAR models are mathematical models that develop correlations between chemical structures and biological activities[125]. QSAR-based methods, enhanced with ML, include decision trees, RFs, SVM, XGBoost, and DL[126]. They have been shown to be more predictive and support the identification of potential compounds. Recent examples of such applications include the Cloud 3D-QSAR model for monoamine oxidase B inhibitors and the XGBoost-driven QSAR model for AKT inhibitors[127].

ADMET profiling. Accurate prediction of Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) can be used to determine the best compounds with respect to pharmacokinetic and toxicity properties[128]. Artificial intelligence (AI) models, such as DeepTox, XGBoost-based ADMET predictors, and convolutional neural network (CNN)-based cytotoxicity predictors, can improve early-stage toxicity screening of small molecules[129]. Accessible resources for in silico ADMET evaluation include platforms such as ADMETlab, SwissADME, admetSAR, and ProTox-II[130].

In summary, AI plays a crucial role in enhancing the speed, scalability, and predictive accuracy of structure-based and ligand-based virtual screening. Combined with docking, MD simulation, QSAR modelling, and ADMET prediction, AI can be used to perform more effective hit identification and lead optimization. With the combination of pattern recognition, structural prediction, and data-driven decision-making frameworks, AI systems are becoming an increasingly important part of a lean end-to-end virtual screening pipeline that accelerates the in-silico discovery-to-experimental validation process.

De novo drug design using generative AI

Generative AI uses DL techniques to create new information, such as text, images, and chemical structures, based on the patterns learned from the training data[131]. Generative AI has also been used in drug discovery, where it can be used to generate drug structures that are completely novel, with certain therapeutic properties, a process called de novo drug design. Because it is estimated that the chemical space contains 10^{23} to 10^{60} possible molecules, current computational screening is not suitable for exhaustively covering it[132,133]. De novo design overcomes this drawback by synthesizing new compounds without relying on existing templates, thereby increasing the chemical space available for drug discovery[134]. As with virtual screening, de novo design can be performed using either structure-based or ligand-based methodologies[135].

The application of neural networks and reinforcement learning (deep reinforcement learning) has taken center stage in generative chemistry because they can process data at a high dimensionality in molecules and identify complex structures[136]. Molecular generation has been successfully achieved using DRL models, including VAEs, CNNs, GANs, and RNNs[21]. For example, Li et al. used an RNN architecture to create new SARS-CoV-2 inhibitors, thus demonstrating the potential of sequence-based generative systems[137].

Variational Autoencoders are unsupervised deep neural networks that learn the latent probability distributions of molecular structures[138]. The encoder is one of the components that converts inputs, such as fingerprints, graphs, or even SMILES strings, to a continuous latent space. Subsequently, the decoder is used to sample this space to produce novel molecular entities. New methods involve hybrid variational autoencoder (VAE) RL systems and generative adversarial network (GAN) designs, both of which are more diverse and chemically valid[139]. It is important to note that PaccMannRA uses the transcriptomic profiles of a given cancer in the VAE-RL pipeline, thus producing compounds that match the disease-specific gene expression rates. This represents the new importance of multi-omics-informed generative design.

RNNs, which produce sequences based on learned time dependencies, have been used to synthesize molecules atom-by-atom using the SMILES notation[140]. Mienye et al. used an RNN to design new JAK2 inhibitors and trained an RNN on ChEMBL to generate new 5-HT2A receptor ligands, and more than 95% of the generated molecules were predicted to be active[141].

The generative platform DeepScaffold, which relies on CNNs, creates novel scaffolds by sequentially adding atoms and bonds in accordance with chemically valid rules; thus, it is a powerful scaffold-hopping tool[142].

GANs are designed using a generator-discriminator architecture to generate realistic molecules[143]. GANs are particularly useful for de novo drug design because they can directly and explicitly handle molecular graphs and SMILES strings without requiring a structural template. GAN-based methods have been used to generate drug-like molecules with extremely high affinity, and more recently, transcriptomic-based GANs have been used to create compounds that specifically target various cancer subtypes[144].

New generative processes are better formulated with synthesis-aware modules to ensure chemical viability. AI-powered synthesis prediction systems (such as ASKCOS and IBM RXN in the case of Chemistry) can be integrated into generative method pipelines to assess the synthetic availability of compounds[145]. In addition, when augmented with generative models and AI-assisted docking, MD simulations, ADMET prediction, and closed-loop optimization, in which generated molecules are optimized based on predicted binding affinities and physicochemical properties, can be performed[146].

Generative AI represents a wide range of architectures, each with its own advantages for de novo molecular design. They can be searched across large chemical spaces, not just in existing libraries, using VAEs, GANs, RNNs, transformer-based models, and emerging diffusion models[136]. Even more recent hybrid and transformer-based models, such as MolGPT and Chemformer, combine structural learning with the capability to generate syntactically coherent and pharmacologically relevant SMILES strings[147]. Such innovations are supplemented by generative pipelines biologically

Table 3 | Comparative summary of key generative ai models used in oncology drug design

Generative model class	Core algorithmic framework	Input representation	Primary applications in oncology drug discovery	Key limitations	References
VAEs	Latent-space encoding/decoding with probabilistic sampling	SMILES strings, molecular fingerprints, graph embeddings	Generation of drug-like molecules; property optimization; multi-omics–conditioned design (e.g., *PaccMannRL*)	Limited chemical validity; latent space may lack interpretability; requires post-generation filtering for synthesis feasibility	138,185
GANs	Generator–discriminator adversarial training	SMILES or molecular graphs	Scaffold innovation; generation of structurally diverse anticancer molecules	Training instability; mode collapse; difficulty incorporating synthetic constraints	186,187
RNNs	Sequence modeling with LSTM/GRU recurrent units	SMILES sequences	Atom-by-atom molecule generation; creation of focused anticancer libraries	Limited ability to capture molecular topology; biased toward training distribution; struggles with long-range structural dependencies	188,189
Transformer-Based Models (e.g., MolGPT, Chemformer)	Self-attention–based sequence modeling; encoder–decoder architectures	SMILES strings, tokenized molecular descriptors	High-fidelity SMILES generation; conditional molecule synthesis; multitask molecular design	High computational cost; requires large datasets; interpretability challenges	190,191
VAE–GAN and Hybrid Models	Combined latent-space learning plus adversarial refinement	SMILES or graph-based encodings	Improved molecular diversity; structural validity enhancement; generative reinforcement learning integration	Complex training; harder to tune; limited validation pipelines	192,193
GraphVAEs, GraphGANs, GraphNVP	Graph neural networks with node/edge generation	Molecular graphs	Topology-aware generation; reactivity-aware anticancer compound design; scaffold hopping	High computational overhead; enforcing chemical rules is challenging; interpretability remains limited	194,195
Diffusion Models	Stochastic denoising processes generating molecules from noise	Graphs, 3D conformers, SMILES	High-quality, diverse molecular generation; shape-conditioned molecule design	Emerging methods; limited oncology-specific benchmarks; synthetic feasibility not inherently encoded	196,197
RL-Driven Generative Models	Policy-gradient optimization over molecular space	SMILES or graph encodings	Target-specific compound optimization; multi-objective scoring; integration with docking and transcriptomic profiles	Reward function design is nontrivial; may generate non-synthesizable molecules; validation requires extensive post-processing	148,198,199

conditioned by integrating disease-specific molecular signatures. An attempt of this sort is PaccMannRL, which trains a VAE-based reinforcement learning model using a transcriptomic database to produce anticancer molecules according to tumor-specific patterns of gene expression[148].

Table 3 provides an overview of the key classes of generative models, their inputs, algorithms, and key areas of application in oncology. Although they can quickly be made more useful, generative models have numerous limitations: they cannot provide an interpretable representation of latent molecular representations, do not fully integrate synthetic feasibility, and do not exist as standardized evaluation pipelines to tie in silico molecular generation with downstream docking, MD simulations, retrosynthetic analysis, and experimental validation.

Generative AI provides a general-purpose platform for searching large chemical spaces and generating new molecules with optimized therapeutic properties. Nevertheless, several limitations exist, including issues with model interpretability, synthetic feasibility, and the need for standardized validation pipelines to effectively bridge the gap between in silico design and experimental and clinical results[149].

Challenges and outlook

Artificial intelligence has significantly accelerated the initial phases of cancer drug development by enabling rapid compound screening, minimizing labor and costs, and maximizing hit identification[36,150]. Nonetheless, it has several limitations that prevent its complete translational use. These problems may be grouped into the following broad categories: data quality and bias issues, model interpretability issues, computational and infrastructural requirements, and ethical and regulatory issues.

Data quality, bias, and generalization

The quality and availability of training data greatly determine the performance of AI systems[27]. Objects in drug discovery may be incomplete, proprietary, or fragmented, thus restricting the strength of the models[151]. Not all pharmacological data are openly available, which limits reproducibility and validation[152,153].

Artificial intelligence systems are vulnerable to biases embedded in the training data, especially when datasets are unbalanced, biased towards a given disease, or poorly representative of rare cancers or poorly studied patient groups[154]. Such biases may extend to predictions, thereby compromising the system's accuracy. These issues can be reduced using strategies such as training across multiple datasets, running numerous trials, and using ensemble models.

A related issue is overfitting, where algorithms learn the noise in the data rather than the generalizable patterns. This reduces their ability to accurately work with new, real-life information. Thus, AI-based predictions must be experimentally verified.

Model opacity and explainable AI (XAI)

The fact that they are black boxes is one of the most significant problems when it comes to implementing DL models, and therefore neural networks. Decision-making in these models within the organization is often opaque, and it is difficult to determine why a model selects a specific target, forecasts that an interaction will occur, or even why it chooses to focus on a molecule. This is not clear in the healthcare sector, where model accountability is the most significant concern[155,156].

Explainable AI (XAI) seeks to address this issue by developing comprehensible frameworks and visualization techniques to explain the prediction process[157,158]. Recent oncology-relevant applications include saliency mapping for drug-target interaction models and attention-based interpretability of graph neural networks. These strategies will build confidence, increase regulatory acceptance, and reduce deployments[159,160].

XAI has become an integral part of oncology AI-driven systems, and transparency is an inherent characteristic of these systems that encourages clinical plausibility and regulatory adherence[161,162]. DL architectures can also provide black-box predictions, making them difficult to interpret, particularly for drug target identification, drug-target interactions, or ranking de

novo generated compounds[163,164]. XAI techniques aim to understand the internal reasoning of such models by determining the molecular, atomic, or structural characteristics that influence predictions. In oncology drug discovery, saliency-based visualization methods have been applied to highlight pharmacophoric substructures that can describe the anticipated anticancer activity and be verified using existing mechanistic knowledge[165,166]. Similarly, attention-based GNNs are interpretable models that can identify essential residues or graph motifs mediating drug-protein interactions by assigning node-based significance scores[167]. The latest uses include gradient-based attribution models that can rediscover the specificity of kinase inhibitors and attention-enhanced GNN models that can identify functionally relevant substituents that result in cytotoxic outcomes. XAI also enhances interpretability to increase scientific confidence, facilitates hypothesis generation by conducting further testing, and enables more informed decision-making, thereby enhancing the translational reliability of AI-based oncology drug discovery[163,164].

Computational and infrastructural constraints

Implementing AI in drug discovery requires substantial computational resources, specialized hardware, and personnel to operate the systems[168]. Many research centres and clinical laboratories lack systems to implement large-scale AI tools regularly.

Moreover, AI models still struggle to capture the complexity of biological systems, such as tumour heterogeneity, temporal dynamics, microenvironmental interactions, and adaptive responses. Even with the development of structural prediction, generative design, and MD simulations, it is challenging to predict biologically complex phenotypes with precision.

Ethical, privacy, and regulatory considerations

AI-based drug discovery may be based on delicate genomic, transcriptomic, and clinical data from patients. Data security and the privacy of patient information are essential, especially given the higher level of technology than the current regulatory systems[168]. Concerns regarding algorithmic transparency, compliance with the FDA, responsibility in clinical decision-making, and data-sharing guidelines have also been raised[169]. Regulatory bodies have yet to create standardized guidelines, and the effectiveness of an algorithm and its impact on the real world must be evaluated.

Clinical translation and real-world impact

Despite of certain challenges, clinically important outcomes of AI have already been achieved. A number of AI-designed or AI-prioritised compounds have been progressed to the human clinical trials, which implies greater target specificity, reduced discovery times, and clinical success.

Nevertheless, AI is no longer a theoretical idea but is delivering practical outcomes. In the last five years, several AI-developed pharmaceuticals have moved to clinical development[170,171]. Small-molecule candidates, including REC-2282, BPM31510, and RLY-4008, have been granted FDA Fast Track designation to expedite their preclinical development. These successes highlight the feasibility of AI-assisted drug discovery and provide a significant impetus for its further adoption.

Future directions

The future of AI-based drug discovery will rely on generating more multimodal data, improving model interpretability, and aligning with experimental feedback systems. Advances in foundational models, multimodal learning, and self-supervised training are expected to improve generalization and reduce reliance on large datasets. Moreover, a combination of AI, automated synthesis platforms, digital twins, robotic screening, and quantum computing would help establish closed-loop adaptive pipelines that actively optimize molecular designs based on real-time biological responses. This merging is bound to close the gap between in silico predictions and clinical translations. The convergence of AI with quantum computing, generative models, and digital twins is poised to enable closed-loop, adaptive drug discovery pipelines, bridging the gap between in silico prediction and clinical translation.

Limitations of this review

The main areas covered in this review are early discovery applications, such as target identification, virtual screening, and de novo design, but it does not cover downstream optimization, formulation, or post-marketing analytics. Some of the tools mentioned are more likely to change or be replaced by newer architectures due to the rapid advancement of AI technology.

Conclusion

Artificial intelligence has rapidly become an essential driver of innovation in the discovery of oncological drugs, significantly accelerating the process of identifying targets, enabling screening large numbers of virtual targets, and aiding de novo molecular design. The combination of multi-omics data, structural prediction, and interaction modelling using ML and generative frameworks enables AI to achieve unprecedented speed and accuracy compared to conventional methodologies. Although there are issues with data quality, model interpretability, computational resources, and regulatory acceptability, recent clinical developments suggest that AI-designed therapeutics are no longer in silico predictions but are now being put into practice as oncology pipelines. Further breakthroughs in multimodal foundation models, XAI, automated synthesis, and closed-loop discovery systems are expected to help bridge the gap between computational design and clinical translation, making AI a revolutionary element in the next generation of cancer treatments.

Received: 20 October 2025; Accepted: 25 January 2026;
Published online: 14 February 2026

References

1. Gupta, R. et al. Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Mol. Divers.* **25**, 1315–1360 (2021).
2. Tsou, L. K. et al. Comparative study between deep learning and QSAR classifications for TNBC inhibitors and novel GPCR agonist discovery. *Sci. Rep.* **10**, https://doi.org/10.1038/s41598-020-73681-1 (2020).
3. Grabowski, H. & Long, G. Post-approval indications and clinical trials for cardiovascular drugs: some implications of the US Inflation Reduction Act. *J. Med. Econ.* **27**, 463–472 (2024).
4. Barbolosi, D., André, N., Barlési, F., Lacarelle, B. & Ciccolini, J. Computational oncology–mathematical modelling of drug regimens for precision medicine. *Nat. Rev. Clin. Oncol.* **13**, 242–254 (2015).
5. Bhange, M. & Telange, D. Convergence of nanotechnology and artificial intelligence in the fight against liver cancer: a comprehensive review. *Discover Oncol.* **16**, https://doi.org/10.1007/s12672-025-01821-y (2025).
6. Biankin, A. V., Piantadosi, S. & Hollingsworth, S. J. Patient-centric trials for therapeutic development in precision oncology. *Nature* **526**, 361–370 (2015).
7. Wang, K., Huang, Y., Wang, Y., You, Q. & Wang, L. Recent advances from computer-aided drug design to artificial intelligence drug design. *RSC Med.Chem.* **15**, 3978–4000 (2024).
8. Rawat, S. et al. Drug repositioning using computer-aided drug design (CADD. *Curr. Pharm. Biotechnol.* **25**, 301–312 (2024).
9. Afrose, N., Chakraborty, R., Bhowmick, M., Bhowmick, P. & Hazra, A. AI-driven drug discovery and development. *Drug Discov Today*. **26**, 80–93 (2024).
10. Lee, J. W., Yoon, S., Maria-Solano, M. A., Choi, S. & Vu, T. N. L. Big data and artificial intelligence (AI) methodologies for computer-aided drug design (CADD. *Biochem. Soc. Trans.* **50**, 241–252 (2022).
11. Udegbe, F., Ekesiobi, C., Ebulue, C. & Ebulue, O. Machine learning in drug discovery: a critical review of applications and challenges. *Comput. Sci. IT Res. J.* **5**, 892–902 (2024).

12. Zhang, X. C. et al. MG-BERT: leveraging unsupervised atomic representation learning for molecular property prediction. *Brief Bioinformatics* **22**, https://doi.org/10.1093/bib/bbab152 (2021).
13. Striuk, O. & Kondratenko, Y. Generative adversarial neural networks and deep learning: successful cases and advanced approaches. *Int. J. Comput* **20**, 339–349 (2021).
14. Singh, A., Eising, C., Ven, P. & Denny, P. Dynamic filter application in graph convolutional networks for enhanced spectral feature analysis and class discrimination in medical imaging. *IEEE Access* **12**, 113259–113274 (2024).
15. Cheung, M. & Moura, J. M. F. Graph Neural Networks for COVID-19 Drug Discovery. In *Proc. IEEE International Conference on Big Data (Big Data)* 5646–5648 (IEEE, 2020).
16. Hanassab, S. et al. The prospect of artificial intelligence to personalize assisted reproductive technology. *npj Digit. Med.* **7**, 55 (2024).
17. Liu, X., Xue, Z., Luo, M., Ke, B. & Lv, J. Anesthetic drug discovery with computer-aided drug design and machine learning. *Anesthesiol. Perioper. Sci.* **2**, https://doi.org/10.1007/s44254-023-00047-x (2024).
18. Wong, C. F. & McCammon, J. A. Protein flexibility and computer-aided drug design. *Annu. Rev. Pharmacol. Toxicol.* **43**, 31–45 (2002).
19. Pasrija, P., Jha, P., Chopra, M., Upadhyaya, P. & Khan, M. S. Machine learning and artificial intelligence: a paradigm shift in big data-driven drug design and discovery. *Curr. Top. Med. Chem.* **22**, 1692–1727 (2022).
20. Nayarisseri, A. et al. Artificial Intelligence. Big data and machine learning approaches in precision medicine & drug discovery. *Curr. Drug Targets* **22**, 631–655 (2021).
21. Cesaro, A., Bagheri, M., Torres, M., Wan, F. & Fuente-Nunez, C. Deep learning tools to accelerate antibiotic discovery. *Expert Opin Drug Discov.* **18**, 1245–1257 (2023).
22. Nguyen, A. T. N. et al. The application of artificial intelligence to accelerate G protein-coupled receptor drug discovery. *Br. J. Pharmacol.* **181**, 2371–2384 (2023).
23. Mouchlis, V. D. et al. Advances in de novo drug design: from conventional to machine learning methods. *Int. J. Mol. Sci.* **22**, https://doi.org/10.3390/ijms22041676 (2021).
24. Lin, E., Lin, C. H. & Lane, H. Y. De novo peptide and protein design using generative adversarial networks: an update. *J. Chem. Inf. Model.* **62**, 761–774 (2022).
25. Fu, C. & Chen, Q. The future of pharmaceuticals: artificial intelligence in drug discovery and development. *J. Pharm. Anal.* **15**, 101248. https://doi.org/10.1016/j.jpha.2025.101248 (2025).
26. Kant, S., Deepika, D. & Roy, S. Artificial intelligence in drug discovery and development: transforming challenges into opportunities. *Discov. Pharm. Sci.* **1**, https://doi.org/10.1007/s44395-025-00007-3 (2025).
27. Ocana, A. et al. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. *Biomarker Res.* **13**, https://doi.org/10.1186/s40364-025-00758-2 (2025).
28. Pammolli, F. et al. The endless frontier? The recent increase of R&D productivity in pharmaceuticals. *J. Translat. Med.* **18**, https://doi.org/10.1186/s12967-020-02313-z (2020).
29. Dhudum, R., Pawar, A. & Ganeshpurkar, A. Revolutionizing drug discovery: a comprehensive review of AI applications. *Drugs Drug Candidates* **3**, 148–171 (2024).
30. Shaki, F., Jahani, D., Amilrkhanloo, M. & Chahrdori, M. Artificial intelligence in pharmaceuticals: exploring applications and legal challenges. *Pharm. Biomed. Res.* **10**, 1–10 (2024).
31. Mukaidaisi, M., Grantham, K., Vu, A., Li, Y. & Tchagang, A. Multi-objective drug design based on graph-fragment molecular representation and deep evolutionary learning. *Front. Pharmacol.* **13**, https://doi.org/10.3389/fphar.2022.920747 (2022).
32. Khan, M. K. et al. The recent advances in the approach of artificial intelligence (AI) towards drug discovery. *Front. Chem.* **12**, https://doi.org/10.3389/fchem.2024.1408740 (2024).
33. Li, K., Du, Y., Li, L. & Wei, D. Q. Bioinformatics approaches for anti-cancer drug discovery. *Curr. Drug Targets* **21**, 3–17 (2019).
34. Doherty, G. J. et al. Cancer treatment in the genomic era. *Annu. Rev. Biochem.* **88**, 247–280 (2019).
35. Kumar, B., Singh, S., Kumar, V. & Skvortsova, I. Promising targets in anti-cancer drug development: recent updates. *Curr. Med. Chem.* **24**, https://doi.org/10.2174/0929867324666170331123648 (2018).
36. Garg, P. et al. Artificial intelligence–driven computational approaches in the development of anticancer drugs. *Cancers* **16**, 3884 (2024).
37. Hachem, S. et al. Contemporary update on clinical and experimental prostate cancer biomarkers: a multi-omics-focused approach to detection and risk stratification. *Biology* **13**, https://doi.org/10.3390/biology13100762 (2024).
38. Arya, K. R. et al. Multitarget-directed multiple ligands in anti-VEGF resistant glioblastoma therapeutics. https://doi.org/10.47852/bonviewMEDIN52023816 (2025).
39. Akbani, R. et al. A pan-cancer proteomic perspective on The Cancer Genome Atlas. *Nat. Commun.* **5**, https://doi.org/10.1038/ncomms4887 (2014).
40. Mottaghi-Dastjerdi, N. & Soltany-Rezaee-Rad, M. Advancements and applications of artificial intelligence in pharmaceutical sciences: a comprehensive review. *Iran. J. Pharm. Res. IJPR* **23**, https://doi.org/10.5812/ijpr-150510 (2024).
41. Kamya, P. et al. PandaOmics: an AI-driven platform for therapeutic target and biomarker discovery. *J. Chem. Inf. Model.* **64**, 3961–3969 (2024).
42. Raufaste-Cazavieille, V., Droit, A. & Santiago, R. Multi-omics analysis: paving the path toward achieving precision medicine in cancer treatment and immuno-oncology. *Front. Mol. Biosci.* **9**, https://doi.org/10.3389/fmolb.2022.962743 (2022).
43. Kamya, P. et al. PandaOmics: an AI-driven platform for therapeutictarget and biomarker discovery. *J. Chem. Inf. Model.* **64**, 3961–3969 (2024).
44. Ivanenkov, Y. A. et al. Chemistry42: an AI-driven platform for molecular design and optimization. *J. Chem. Inf. Model.* **63**, 695–701 (2023).
45. Patel, M. N., Halling-Brown, M. D., Al-Lazikani, B., Tym, J. E. & Workman, P. Objective assessment of cancer genes for drug discovery. *Nat. Rev. Drug Discov.* **12**, 35–50 (2012).
46. Migliozzi, S. et al. Integrative multi-omics networks identify PKCδ and DNA-PK as master kinases of glioblastoma subtypes and guide targeted cancer therapy. *Nat. Cancer* **4**, 181–202 (2023).
47. Tang, X. et al. A survey of generative AI for de novo drug design: new frontiers in molecule and protein generation. *Briefings Bioinformatics* **25**, https://doi.org/10.1093/bib/bbae338 (2024).
48. Bilal, H. et al. An intelligent approach for early and accurate predication of cardiac disease using hybrid artificial intelligence techniques. https://doi.org/10.3390/bioengineering11121290 (2024).
49. Karimi, M., Wang, Z., Shen, Y. & Wu, D. DeepAffinity: interpretable deep learning of compound-protein affinity through unified recurrent and convolutional neural networks. *Bioinformatics* **35**, 3329–3338 (2019).
50. Lilhore, U. K. et al. ProtienCNN-BLSTM: an efficient deep neural network with amino acid embedding-based model of protein sequence classification and biological analysis. *Comput. Intelligence* **40**, https://doi.org/10.1111/coin.12696 (2024).
51. Bull, S. C. & Doig, A. J. Properties of protein drug target classes. *PLoS ONE* **10**, https://doi.org/10.1371/journal.pone.0117955 (2015).

52. Haider, S., Ghosh, S., Pal, R. & Rahman, R. A copula based approach for design of multivariate random forests for drug sensitivity prediction. *PLOS ONE* **10**, 0144490 (2015).
53. Akter, S. AI-driven precision medicine: transforming personalized cancer treatment. *J. AI Power. Med. Innov.* **2**, 10–21 (2024).
54. Masso, M. & Vaisman, I. I. AUTO-MUTE 2.0: a portable framework with enhanced capabilities for predicting protein functional consequences upon mutation. *Adv. Bioinformatics* **2014**, 1–7 (2014).
55. Zhai, D. et al. Small-molecule targeting AMPA-mediated excitotoxicity has therapeutic effects in mouse models for multiple sclerosis. *Sci. Adv.* **9**, https://doi.org/10.1126/sciadv.adj6187 (2023).
56. Song, K. et al. AlloDriver: a method for the identification and analysis of cancer driver targets. *Nucleic Acids Res.* **47**, 315–321 (2019).
57. Larochelle, J. R. et al. Structural reorganization of SHP2 by oncogenic mutations and implications for oncoprotein resistance to allosteric inhibition. *Nat. Commun.* **9**, https://doi.org/10.1038/s41467-018-06823-9 (2018).
58. Deng, X., Das, S., Valdez, K., Camphausen, K. & Shankavaram, U. SL-BioDP: Multi-cancer interactive tool for prediction of synthetic lethality and response to cancer treatment. *Cancers* **11**, https://doi.org/10.3390/cancers11111682 (2019).
59. Das, S., Shankavaram, U., Deng, X. & Camphausen, K. DiscoverSL: an R package for multi-omic data driven prediction of synthetic lethality in cancers. *Bioinformatics* **35**, 701–702 (2018).
60. Benfatto, S. et al. Uncovering cancer vulnerabilities by machine learning prediction of synthetic lethality. *Mol. Cancer* **20**, https://doi.org/10.1186/s12943-021-01405-8 (2021).
61. Huang, J., Lu, F., Wu, M., Ou-Yang, L. & Zhu, Z. Predicting synthetic lethal interactions in human cancers using graph regularized self-representative matrix factorization. *BMC Bioinformatics* **20**, https://doi.org/10.1186/s12859-019-3197-3 (2019).
62. Wang, Z. et al. BatmanNet: bi-branch masked graph transformer autoencoder for molecular representation. *Brief. Bioinformatics* **25**, https://doi.org/10.1093/bib/bbad400 (2023).
63. Savage, S. R. et al. Pan-cancer proteogenomics expands the landscape of therapeutic targets. *Cell* **187**, 4389–4407 4315 (2024).
64. Kokudeva, M., Vichev, M., Naseva, E., Miteva, D. G. & Velikova, T. Artificial intelligence as a tool in drug discovery and development. *World J. Exp. Med.* **14**, https://doi.org/10.5493/wjem.v14.i3.96042 (2024).
65. Nelen, J. et al. ESSENCE-Dock: a consensus-based approach to enhance virtual screening enrichment in drug discovery. *J. Chem. Inf. Model.* **64**, 1605–1614 (2024).
66. Bielska, E. et al. REVIEW PAPER: Virtual screening strategies in drug design – methods and applications. *BioTechnologia* **92**, 249–264 (2011).
67. Polgar, T. & Keseru, M. Integration of virtual and high throughput screening in lead discovery settings. *Combinat. Chem. High. Throughput Screen.* **14**, 889–897 (2011). & G.
68. Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discov.* **1**, 882–894 (2002).
69. Grotsch, K. et al. Virtual screening of a chemically diverse "Superscaffold. *Library Enables Ligand Discovery for a Key GPCR Target*. *ACS Chem. Biol.* **19**, 866–874 (2024).
70. Yang, Y. et al. D3AI-CoV: a deep learning platform for predicting drug targets and for virtual screening against COVID-19. *Brief. Bioinformatics* **23**, https://doi.org/10.1093/bib/bbac147 (2022).
71. Cournia, Z. et al. Rigorous free energy simulations in virtual screening. *J. Chem. Inf. Model.* **60**, 4153–4169 (2020).
72. Nascimento, L. D. et al. Applications of virtual screening in bioprospecting: facts, shifts, and perspectives to explore the chemo-structural diversity of natural products. *Front. Chem.* **9**, https://doi.org/10.3389/fchem.2021.662688 (2021).
73. Ebalunode, J. O., Liang, J., Ouyang, Z. & Zheng, W. Novel approach to structure-based pharmacophore search using computational geometry and shape matching techniques. *J. Chem. Inf. Model.* **48**, 889–901 (2008).
74. Niinivehmas, S. P., Manivannan, E., Rauhamäki, S., Huuskonen, J. & Pentikäinen, O. T. Identification of estrogen receptor α ligands with virtual screening techniques. *J. Mol. Graph. Model.* **64**, 30–39 (2016).
75. Rudrapal, M. & Chetia, D. Virtual screening, molecular docking and QSAR studies in drug discovery and development programme. *J. Drug Deliv. Ther.* **10**, 225–233 (2020).
76. Potlitz, F., Link, A. & Schulig, L. Advances in the discovery of new chemotypes through ultra-large library docking. *Expert Opin. Drug Discov.* **18**, 303–313 (2023).
77. Jiang, P. et al. Molecular persistent spectral image (Mol-PSI) representation for machine learning models in drug design. *Brief. Bioinformatics* **23**, https://doi.org/10.1093/bib/bbab527 (2021).
78. Kuan, J., Cherkasov, A., Avenido, A., Gentile, F. & Radaeva, M. Keeping pace with the explosive growth of chemical libraries with structure-based virtual screening. *WIREs Comput. Mol. Sci.* **13**, https://doi.org/10.1002/wcms.1678 (2023).
79. Wang, S. et al. MSGNN-DTA: multi-scale topological feature fusion based on graph neural networks for drug-target binding affinity prediction. *Int. J. Mol. Sci.* **24**, 8326 (2023).
80. Isert, C., Atz, K., Riniker, S. & Schneider, G. Exploring protein-ligand binding affinity prediction with electron density-based geometric deep learning. *RSC Adv.* **14**, 4492–4502 (2024).
81. Zhou, G. et al. An artificial intelligence accelerated virtual screening platform for drug discovery. *Nat. Commun.* **15**, https://doi.org/10.1038/s41467-024-52061-7 (2024).
82. Meli, R., Morris, G. M. & Biggin, P. C. Scoring functions for protein-ligand binding affinity prediction using structure-based deep learning. *Rev. Front. Bioinformatics* **2**, https://doi.org/10.3389/fbinf.2022.885983 (2022).
83. Bekker, G. J., Kamiya, N., Fukuda, I., Fukunishi, Y. & Higo, J. Cryptic-site binding mechanism of medium-sized Bcl-xL inhibiting compounds elucidated by McMD-based dynamic docking simulations. *Sci. Rep.* **11**, https://doi.org/10.1038/s41598-021-84488-z (2021).
84. Vidal-Limon, A., Aguilar-Toalá, J. E. & Liceaga, A. M. Integration of molecular docking analysis and molecular dynamics simulations for studying food proteins and bioactive peptides. *J. Agric. Food Chem.* **70**, 934–943 (2022).
85. Beglov, D. et al. Exploring the structural origins of cryptic sites on proteins. *Proc. Natl. Acad. Sci. USA* **115**, https://doi.org/10.1073/pnas.1711490115 (2018).
86. Naqvi, A. A. T., Hassan, M. I., Mohammad, T. & Hasan, G. M. Advancements in docking and molecular dynamics simulations towards ligand-receptor interactions and structure-function relationships. *Curr. Top. Med. Chem.* **18**, 1755–1768 (2018).
87. Gu, S. et al. Can molecular dynamics simulations improve predictions of protein-ligand binding affinity with machine learning? *Brief. Bioinformatics* **24**, https://doi.org/10.1093/bib/bbad008 (2023).
88. Bentham Science Publisher, B. S. P. Docking and scoring - theoretically easy, practically impossible? *Curr. Med.Chem.* **13**, 2995–3003 (2006).
89. Díaz-Holguín, A. et al. AlphaFold accelerated discovery of psychotropic agonists targeting the trace amine-associated receptor 1. *Sci. Adv.* **10**, https://doi.org/10.1126/sciadv.adn1524 (2024).

90. Glukhov, E. et al. *Phospho-Tune: Enhanced Structural Modeling of Phosphorylated Protein Interactions* (Cold Spring Harbour Laboratory, 2024).
91. Liu, J., Neupane, P. & Cheng, J. in *and AlphaFold3-Based Protein Complex Structure Prediction With MULTICOM4 in CASP16*. *Proteins* (2025).
92. Krokidis, M. G. et al. AlphaFold3: an overview of applications and performance insights. *Int. J. Mol. Sci.* **26**, 3671 (2025).
93. Gheidari, D., Mehrdad, M. & Bayat, M. Synthesis, docking, MD simulation, ADMET, drug likeness, and DFT studies of novel furo[2,3-b]indol-3a-ol as promising Cyclin-dependent kinase 2 inhibitors. *Sci. Rep.* **14**, https://doi.org/10.1038/s41598-024-53514-1 (2024).
94. Kaveh, S., Mani-Varnosfaderani, A. & Neiband, M. S. Deriving general structure–activity/selectivity relationship patterns for different subfamilies of cyclin-dependent kinase inhibitors using machine learning methods. *Sci. Rep.* **14**, https://doi.org/10.1038/s41598-024-66173-z (2024).
95. Sheik Amamuddy, O. et al. Integrated computational approaches and tools forallosteric drug discovery. *Int. J. Mol. Sci.* **21**, 847 (2020).
96. Aderinwale, T. et al. Real-time structure search and structure classification for AlphaFold protein models. *Commun. Biol.* **5**, https://doi.org/10.1038/s42003-022-03261-8 (2022).
97. Wu, N., Strömich, L. & Yaliraki, S. N. Prediction of allosteric sites and signaling: insights from benchmarking datasets. *Patterns* **3**, 100408 (2021).
98. Ruiz-Carmona, S. et al. Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat. Chem.* **9**, 201–206 (2016).
99. Sun, H. & Scott, D. O. Structure-based drug metabolism predictions for drug design. *Chem. Biol. Drug Des.* **75**, 3–17 (2009).
100. Chen, L. et al. From laptop to benchtop to bedside: structure-based drug design on protein targets. *Curr. Pharm. Des.* **18**, 1217–1239 (2012).
101. Ivanova, L. & Karelson, M. The impact of software used and the type of target protein on molecular docking accuracy. *Molecules* **27**, 9041 (2022).
102. Pinzi, L. & Rastelli, G. Molecular docking: shifting paradigms in drug discovery. *Int. J. Mol. Sci.* **20**, 4331 (2019).
103. Aboalroub, A. A. Virtual screening and molecular docking characterization of isoxazole-based small molecules as potential Hsp90 inhibitors: an in silico insight. *Medinformatics* **2**, 154–166 (2025).
104. Luttens, A. et al. Rapid traversal of vast chemical space using machine learning-guided docking screens. *Nat. Comput. Sci.* **5**, 301–312 (2025).
105. Pason, L. P. & Sotriffer, C. A. Empirical scoring functions for affinity prediction of protein-ligand complexes. *Mol. Inform.* **35**, 541–548 (2016).
106. Morrone, J. A., Luo, H., Huynh, T., Weber, J. K. & Cornell, W. D. Combining docking pose rank and structure with deep learning improves protein-ligand binding mode prediction over a baseline docking approach. *J. Chem. Inf. Model.* **60**, 4170–4179 (2020).
107. Lee, C. et al. GalaxyDock-DL: protein-ligand docking by global optimization and neural network energy. *J. Chem. Theory Comput.* https://doi.org/10.1021/acs.jctc.4c00385 (2024).
108. Buttenschoen, M., Morris, G. M. & Deane, C. M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.* **15**, 3130–3139 (2024).
109. Ghersi, D. & Sanchez, R. Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites. *Proteins* **74**, 417–424 (2008).
110. Arcon, J. P. et al. Molecular dynamics in mixed solvents reveals protein-ligand interactions, improves docking, and allows accurate binding free energy predictions. *J. Chem. Inf. Model.* **57**, 846–863 (2017).
111. Kuzmanic, A., Bowman, G. R., Michel, J., Juarez-Jimenez, J. & Gervasio, F. L. Investigating cryptic binding sites by molecular dynamics simulations. *Acc. Chem. Res.* **53**, 654–661 (2020).
112. Ge, Y., Pande, V., Seierstad, M. J. & Damm-Ganamet, K. L. Exploring the application of sitemap and site finder for focused cryptic pocket identification. *J. Phys. Chem. B* **128**, 6233–6245 (2024).
113. Sabanés Zariquiey, F., Souza, J. V. & Bronowska, A. K. Cosolvent Analysis Toolkit (CAT): a robust hotspot identification platform for cosolvent simulations of proteins to expand the druggable proteome. *Sci. Rep.* **9**, https://doi.org/10.1038/s41598-019-55394-2 (2019).
114. Kashyap, A., Singh, P. K. & Silakari, O. Counting on fragment based drug design approach for drug discovery. *Curr. Top. Med. Chem.* **18**, 2284–2293 (2019).
115. Czub, N. et al. Artificial intelligence-based quantitative structure-property relationship model for predicting human intestinal absorption of compounds with serotonergic activity. *Mol. Pharmaceutics* **20**, 2545–2555 (2023).
116. Qin, T., Zhu, Z., Wang, X. S., Xia, J. & Wu, S. Computational representations of protein–ligand interfaces for structure-based virtual screening. *Expert Opin. Drug Discov.* **16**, 1175–1192 (2021).
117. Muhammed, M. T. & Aki-Yalcin, E. Pharmacophore modeling in drug discovery: methodology and current status. *J. Turkish Chem. Soc. Sect. A Chem.* **8**, 749–762 (2021).
118. Akingbade, T. V. et al. In silico design of novel Artocarpus altilis-derived compounds targeting the c-Myc/Max heterodimer. *Medinformatics* **2**, 181–194 (2025).
119. Fassio, A. V. et al. Prioritizing virtual screening with interpretable interaction fingerprints. *J. Chem. Inf. Model.* **62**, 4300–4318 (2022).
120. Jiang, X., Tan, L. & Zou, Q. DGCL: dual-graph neural networks contrastive learning for molecular property prediction. *Brief. Bioinformatics* **25**, https://doi.org/10.1093/bib/bbae474 (2024).
121. Menke, J. & Koch, O. Using domain-specific fingerprints generated through neural networks to enhance ligand-based virtual screening. *J. Chem. Inf. Model.* **61**, 664–675 (2021).
122. Bathini, R., Fatima, S., Manga, V. & Sivan, S. K. Molecular docking, MM/GBSA and 3D-QSAR studies on EGFR inhibitors. *J. Chem. Sci.* **128**, 1163–1173 (2016).
123. Pérez-Nueno, V. I., Ritchie, D. W., Pettersson, S., Borrell, J. I. & Teixidó, J. Discovery of novel HIV entry inhibitors for the CXCR4 receptor by prospective virtual screening. *J. Chem. Inf. Model.* **49**, 810–823 (2009).
124. Liu, J. et al. Combined pharmacophore modeling, 3D-QSAR and docking studies to identify novel HDAC inhibitors using drug repurposing. *J. Biomol. Struct. Dyn.* **38**, 533–547 (2019).
125. Keyvanpour, M. R. & Shirzad, M. B. An analysis of QSAR research based on machine learning concepts. *Curr. Drug Discov. Technol.* **18**, 17–30 (2020).
126. Babajide Mustapha, I. & Saeed, F. Bioactive molecule prediction using extreme gradient boosting. *Molecules* **21**, 983 (2016).
127. Ghaleb, A., Aouidate, A., Bouachrine, M., Lakhlifi, T. & Sbai, A. In silico exploration of Aryl halides analogues as checkpoint kinase 1 inhibitors by using 3D QSAR, molecular docking study, and ADMET screening. *Adv. Pharm. Bull.* **9**, 84–92 (2019).
128. Halder, A. K. & Cordeiro, M. N. D. S. AKT inhibitors: the road ahead to computational modeling-guided discovery. *Int. J. Mol. Sci.* **22**, 3944 (2021).
129. Xiong, G. et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* **49**, 5–14 (2021).
130. Sá, A. G. C., Ascher, D. B. & Myung, Y. Deep-PK: deep learning for small molecule pharmacokinetic and toxicity prediction. *Nucleic Acids Res.* **52**, 469–475 (2024).
131. Gangwal, A. et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities.

Fron. Pharmacol. **15**, https://doi.org/10.3389/fphar.2024.1331062 (2024).
132. Gupta, A. et al. Generative recurrent networks for de novo drug design. *Mol. Inform.* **37**, 1700111 (2017).
133. Moret, M., Grisoni, F., Friedrich, L., Merk, D. & Schneider, G. Generative molecular design in low data regimes. *Nat. Mach. Intell.* **2**, 171–180 (2020).
134. Loving, K., Alberts, I. & Sherman, W. Computational approaches for fragment-based and de novo design. *Curr. Top. Med. Chem.* **10**, 14–32 (2010).
135. Gummesson Svensson, H., Tyrchan, C., Engkvist, O. & Haghir Chehreghani, M. Utilizing reinforcement learning for de novo drug design. *Mach. Learn.* **113**, 4811–4843 (2024).
136. Sousa, T., Correia, J., Pereira, V. & Rocha, M. Generative deep learning for targeted compound design. *J. Chem. Inf. Model.* **61**, 5343–5361 (2021).
137. Kotsias, P. C. et al. Direct steering of de novo molecular generation with descriptor conditional recurrent neural networks. *Nat. Mach. Intell.* **2**, 254–265 (2020).
138. Lavecchia, A. Navigating the frontier of drug-like chemical space with cutting-edge generative AI models. *Drug Discovery Today* **29**, 104133 (2024).
139. Hong, S. H., Ryu, S., Lim, J. & Kim, W. Y. Molecular generative model based on an adversarially regularized autoencoder. *J. Chem. Inf. Model.* **60**, 29–36 (2019).
140. Monti, M., Fiorentino, J., Gosti, G., Tartaglia, G. G. & Milanetti, E. Prediction of time series gene expression and structural analysis of gene regulatory networks using recurrent neural networks. *Entropy* **24**, 141 (2022).
141. Mienye, I. D., Swart, T. G. & Obaido, G. Recurrent neural networks: a comprehensive review of architectures, variants, and applications. *Information* **15**, 517 (2024).
142. Zheng, S. et al. Deep scaffold hopping with multimodal transformer neural networks. *J. Cheminformatics* **13**, https://doi.org/10.1186/s13321-021-00565-5 (2021).
143. Bian, Y. & Xie, X. Q. Generative chemistry: drug discovery with deep learning generative models. *J. Mol. Model.* **27**, https://doi.org/10.1007/s00894-021-04674-8 (2021).
144. Zhao, L., Wang, J., Pang, L., Liu, Y. & Zhang, J. GANsDTA: predicting drug-target binding affinity using GANs. *Front. Genet.* **10**, https://doi.org/10.3389/fgene.2019.01243 (2020).
145. Baammi, S., El Allali, A. & Daoud, R. Potent VEGFR-2 inhibitors for resistant breast cancer: a comprehensive 3D-QSAR, ADMET, molecular docking and MMPBSA calculation on triazolopyrazine derivatives. *Front. Mol. Biosci.* **10**, https://doi.org/10.3389/fmolb.2023.1288652 (2023).
146. Guo, J. et al. DockStream: a docking wrapper to enhance de novo molecular design. *J. Cheminformatics* **13**, https://doi.org/10.1186/s13321-021-00563-7 (2021).
147. Abate, C., Cavalli, A. & Decherchi, S. Graph neural networks for conditional de novo drug design. *WIREs Comput. Mol. Sci.* **13**, https://doi.org/10.1002/wcms.1651 (2023).
148. Born, J. et al in *Designing Anticancer Drugs From Transcriptomic Data via Reinforcement Learning* 231–233 (Springer, 2020).
149. Huang, D., Yang, C., Li, Z., Li, N. & Jiang, T. Artificial intelligence in lung cancer: current applications, future perspectives, and challenges. *Front. Oncol.* **14**, https://doi.org/10.3389/fonc.2024.1486310 (2024).
150. Noorain, N., Parveen, R., Parveen, B. & Srivastava, V. Artificial intelligence in drug formulation and development: applications and future prospects. *Curr. Drug Metab.* **24**, 622–634 (2023).
151. Hasselgren, C. & Oprea, T. I. Artificial intelligence for drug discovery: are we there yet? *Annu. Rev. Pharmacol. Toxicol.* **64**, 527–550 (2023).
152. Larsson, P. et al. Optimization of cell viability assays to improve replicability and reproducibility of cancer drug sensitivity screens. *Sci. Rep.* **10**, https://doi.org/10.1038/s41598-020-62848-5 (2020).
153. Schaduangrat, N. et al. Towards reproducible computational drug discovery. *J. Cheminformatics* **12**, https://doi.org/10.1186/s13321-020-0408-x (2020).
154. Han, D., Huong, P. & Cheng, S. Enhancing semantic segmentation through reinforced active learning: combating dataset imbalances and bolstering annotation efficiency. *J. Electron. Inf. Syst.* **5**, 45–60 (2023).
155. Karim, M. R. et al. Explainable AI for Bioinformatics: methods, tools and applications. *Brief. Bioinformatics* **24**, https://doi.org/10.1093/bib/bbad236 (2023).
156. Singh, D. S., Pratap Singh, D. D. & Chandra, M. K. Enhancing transparency and interpretability in deep learning models: a comprehensive study on explainable AI techniques. *Int. J. Sci. Res. Eng. Manag.* **08**, 1–13 (2024).
157. Alkhanbouli, R., Matar Abdulla Almadhaani, H., Alhosani, F. & Simsekler, M. C. E. The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions. *BMC Med. Inform. Decision Making* **25**, https://doi.org/10.1186/s12911-025-02944-6 (2025).
158. Başağaoğlu, H. et al. A review on interpretable and explainable artificial intelligence in hydroclimatic applications. *Water* **14**, https://doi.org/10.3390/w14081230 (2022).
159. Korade, D. Unlocking machine learning model decisions: a comparative analysis of LIME and SHAP for enhanced interpretability. *J. Electr. Syst.* **20**, 598–613 (2024).
160. Thalpage, N. Unlocking the black box: explainable artificial intelligence (XAI) for trust and transparency in AI systems. *J. Digit. Art. Human.* **4**, 31–36 (2023).
161. Anang, A. et al. Explainable AI in financial technologies: balancing innovation with regulatory compliance. *Int. J. Sci. Res. Arch.* **13**, 1793–1806 (2024).
162. Metta, C., Giannotti, F., Rinzivillo, S., Beretta, A. & Pellungrini, R. Towards transparent healthcare: advancing local explanation methods in explainable artificial intelligence. *Bioengineering* **11**, 369 (2024).
163. Alizadehsani, R. et al. Explainable artificial intelligence for drug discovery and development: a comprehensive survey. *IEEE Access* **12**, 35796–35812 (2024).
164. Kırboğa, K. K., Abbasi, S. & Küçüksille, E. U. Explainability and white box in drug discovery. *Chem. Biol. Drug Des.* **102**, 217–233 (2023).
165. Jiang, Y. et al. Pharmacophoric-constrained heterogeneous graph transformer model for molecular property prediction. *Commun. Chem.* **6**, https://doi.org/10.1038/s42004-023-00857-x (2023).
166. Yang, Z., Zhong, W., Zhao, L. & Yu-Chian Chen, C. MGraphDTA: deep multiscale graph neural network for explainable drug-target binding affinity prediction. *Chem. Sci.* **13**, 816–833 (2022).
167. Li, S. et al. Structure-aware interactive graph neural networks for the prediction of protein-ligand binding affinity. 975–985 (2021).
168. Singhal, S. Data privacy, compliance, and security including AI ML. Healthcare 111–126 (2024).
169. Koo, T. H., Zakaria, A. D., Ng, J. K. & Leong, X. B. Systematic review of the application of artificial intelligence in healthcare and nursing care. *Malays. J. Med. Sci. MJMS* **31**, 135–142 (2024).
170. Kalra, B., Khirasaria, R. & Batta, A. Trends in FDA drug approvals over last 2 decades: An observational study. *J. Fam. Med. Prim. Care* **9**, 105 (2020).
171. Vora, L. K. et al. Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics* **15**, https://doi.org/10.3390/pharmaceutics15071916 (2023).
172. O'Connor, L. M., O'Connor, B. A., Lim, S. B., Zeng, J. & Lo, C. H. Integrative multi-omics and systems bioinformatics in translational

neuroscience: a data mining perspective. *J. Pharm. Anal.* **13**, 836–850 (2023).
173. Alghushairy, O. et al. Machine learning-based model for accurate identification of druggable proteins using light extreme gradient boosting. *J. Biomol. Struct. Dyn.* **42**, 12330–12341 (2023).
174. Haza, K. Z. et al. RAS-inhibiting biologics identify and probe druggable pockets, including an SII- u03b13 allosteric site. *Nat. Commun.* **12**, https://doi.org/10.1038/s41467-021-24316-0 (2021).
175. Meng, L., Ye, Z., Yang, Y. & Zhao, H. DeepMCGCN: multi-channel deep graph neural networks. *Int. J. Comput. Intell. Syst.* **17**, https://doi.org/10.1007/s44196-024-00432-9 (2024).
176. Bhalla, S. & Laganà, A. Artificial intelligence for precision oncology. *Adv. Exp. Med. Biol.* **1361**, 249–268 (2022).
177. Eckhart, L., Lenhof, K., Rolli, L. M. & Lenhof, H. P. A comprehensive benchmarking of machine learning algorithms and dimensionality reduction methods for drug sensitivity prediction. *Brief. Bioinformatics* **25**, https://doi.org/10.1093/bib/bbae242 (2024).
178. McDonagh, E. M. et al. Human genetics and genomics for drug target identification and prioritization: open targets' perspective. *Annu. Rev. Biomed. Data Sci.* **7**, 59–81 (2024).
179. Kim, M., Oh, I. & Ahn, J. An improved method for prediction of cancer prognosis by network learning. *Genes* **9**, 478 (2018).
180. Dezső, Z. & Ceccarelli, M. Machine learning prediction of oncology drug targets based on protein and network properties. *BMC Bioinformatics* **21**, https://doi.org/10.1186/s12859-020-3442-9 (2020).
181. Raies, A. et al. DrugnomeAI is an ensemble machine-learning framework for predicting druggability of candidate drug targets. *Commun. Biol.* **5**, https://doi.org/10.1038/s42003-022-04245-4 (2022).
182. Nussinov, R., Jang, H. & Tsai, C. Autoinhibition can identify rare driver mutations and advise pharmacology. *FASEB J.* **34**, 16–29 (2019).
183. Song, Q. et al. DeepAlloDriver: a deep learning-based strategy to predict cancer driver mutations. *Nucleic Acids Res.* **51**, 129–133 (2023).
184. Park, N., Dong, X. L., Kan, A., Zhao, T. & Faloutsos, C. *Estimating Node Importance In Knowledge Graphs Using Graph Neural Networks* 596–606 (Association for Computing Machinery, 2019).
185. Li, P. et al. Improving drug response prediction via integrating gene relationships with deep learning. *Brief. Bioinformatics* **25**, https://doi.org/10.1093/bib/bbae153 (2024).
186. Purwono, P., Wulandari, A. N. E., Salah, W. A. & Ma'Arif, A. Understanding Generative Adversarial Networks (GANs): A Review. *Control Syst. Optim. Lett.* **3**, 36–45 (2025).
187. Ni, Y., Song, D., Liao, L., Zhang, X. & Wu, H. CAGAN: Consistent Adversarial Training Enhanced GANs https://doi.org/10.24963/ijcai.2018/359 (2018).
188. Chen, Y. et al. Molecular language models: RNNs or transformer? *Brief. Funct. Genomics* **22**, 392–400 (2023).
189. Laurent, C., Zhang, Y., Pereyra, G., Brakel, P. & Bengio, Y. Batch normalized recurrent neural networks. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2657-2661 (IEEE, 2016).
190. Bagal, V., Vinod, P. K., Priyakumar, U. D. & Aggarwal, R. MolGPT: molecular generation using a transformer-decoder model. *J. Chem. Inf. Model.* **62**, 2064–2076 (2021).
191. Ross, J. et al. GP-MoLFormer: a foundation model for molecular generation. https://doi.org/10.48550/arxiv.2405.04912 (2024).
192. Miletic, M. & Sariyar, M. Challenges of using synthetic data generation methods for tabular microdata. *Appl. Sci.* **14**, 5975 (2024).
193. Yu, M. S. et al. A variational autoencoder cascade generative adversarial network for scalable 3D object generation and reconstruction. https://doi.org/10.3390/s24030751 (2024).
194. Cai, L. et al. AEGNN-M:A 3D Graph-Spatial Co-Representation Model for Molecular Property Prediction. *IEEE J. Biomed. Health Inform.* **29**, 1726–1734 (2025).
195. Cao, N. & Kipf, T. *MolGAN: An Implicit Generative Model For Small Molecular Graphs* (Cornell University, 2018).
196. Urbina, F., Lowden, C. T., Culberson, J. C. & Ekins, S. MegaSyn: integrating generative molecular design, automated analog designer, and synthetic viability prediction. *ACS Omega* **7**, 18699–18713 (2022).
197. Xu, C., Zheng, P., Chen, H., Wang, H. & Wang, W. Geometric-facilitated denoising diffusion model for 3D molecule generation. *Proc. AAAI Conf. Artif. Intell.* **38**, 338–346 (2024).
198. Regenwetter, L., Nobari, A. H. & Ahmed, F. Deep generative models in engineering design. *A Review*. *J. Mech. Design* **144**, https://doi.org/10.1115/1.4053859 (2022).
199. Liu, X. et al. DrugEx v2: de novo design of drug molecules by Pareto-based multi-objective reinforcement learning in polypharmacology. *J. Cheminformatics* **13**, https://doi.org/10.1186/s13321-021-00561-9 (2021).

Acknowledgements

This study was funded by the Liaoning Provincial Science and Technology Joint Program (Natural Science Foundation General Program, Grant No. 2024-MSLH-536).

Author contributions

Da Li. Sanbao Shi.: Conceptualization, methodology, writing—original draft; Writing— review. Zhiyu Yu. Peng Xu.: Data curation, Formal analysis; Writing–review and editing. Cheng Zhang.: Supervision, writing–review and editing. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Peng Xu or Cheng Zhang.