HPR
Hospital Practices and Research
http://www.jhpr.ir
Hosp Pract Res. 2026 June;11(2):852-873
doi 10.30491/hpr.2026.549556.1518
Review Article

A Review on Artificial Intelligence Algorithms for Computer-Aided Drug Design Based on Recombinant Proteins in Cancer Therapy

Najmeh Akbari [1], **Mohebali Rahdar** [1*], **Gholamreza Farnoosh** [2], **Ghorbanali Bandani** [3]

[1] Department of Industrial Engineering, University of Sistan and Baluchistan, Zahedan, Iran
[2] Applied Biotechnology Research Center, New Health Technologies Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran
[3] Nabi Akram Hospital, Zahedan, Iran

***Corresponding Author:** Mohebali Rahdar, Ph.D., Department of Industrial Engineering, University of Sistan and Baluchistan, Zahedan, Iran. Tel: +989153410725, Email: m.ali.rahdar@eng.usb.ac.ir

Received September 27, 2025; **Accepted** June 3, 2026; **Online Published** June 30, 2026

Abstract

Background: Artificial Intelligence (AI) and Machine Learning (ML) have transformed computer-aided drug design (CADD) by leveraging big data and advanced algorithms to accelerate drug design. These technologies enhance the exploration of chemical spaces, prediction of drug-target interactions, and development of personalized therapeutics, particularly for complex diseases like cancer.

Objectives: This review aims to evaluate the role of AI and ML in CADD, focusing on their applications in high-throughput screening (HTS), three-dimensional (3D) protein structure prediction, and drug-target identification, while addressing challenges and future prospects.

Methods: A comprehensive analysis of recent literature (2005–2024) was conducted using scientometric tools like VOSviewer to identify trends and keywords in CADD. AI-driven methods, including deep learning frameworks (e.g., TensorFlow, AlphaFold2) and computational techniques (e.g., molecular dynamics simulations), were reviewed for their contributions to drug design.

Results: AI and ML have streamlined HTS, improved 3D protein structure prediction, and enhanced drug-target identification, reducing development timelines and costs. Tools like AlphaFold 2 and QuoteTarget have identified novel drug targets with high accuracy. However, challenges such as data quality, model interpretability, and ethical concerns persist. Interdisciplinary collaboration has driven innovation in personalized therapeutics.

Conclusion: AI and ML have revolutionized CADD, offering efficient and precise solutions for drug design. Overcoming data and ethical challenges through interdisciplinary efforts and advanced algorithms will further enhance the development of targeted therapies, reshaping therapeutic paradigms for complex diseases.

Keywords: Artificial Intelligence, Computer-Aided Drug Design, Machine Learning, Drug Design

1. Background

Drug delivery technology has evolved remarkably, driven by the need to address increasingly complex challenges. The first generation focused on improving the basic physicochemical properties of drugs, such as solubility, stability, and bioavailability. However, biological barriers, including the cellular membrane and the blood-brain barrier, emerged as significant obstacles. This led to the development of second-generation drug delivery systems specifically designed to overcome these biological barriers. Effective and targeted drugs can improve patient recovery and facilitate treatment. By developing potent and specific therapeutics, it is possible to increase survival rates and reduce treatment-related side effects.[1]

Designing therapeutic drugs for cancer represents a highly vital and complex domain in medical and pharmaceutical sciences. These advancements employ mechanisms such as enhanced penetration, targeted delivery to specific tissues or cells, and protection against degradation by the body's defense mechanisms. Ultimately, the aim of these second-generation approaches is to improve therapeutic efficacy while minimizing side effects.[2]

Despite the progress achieved with second-generation drug delivery technologies, it became evident that a comprehensive strategy addressing both physicochemical and biological barriers was necessary. This realization prompted the development of third-generation drug delivery technologies, which aim to integrate solutions for physicochemical and biological challenges. These systems seek to optimize drug properties such as solubility, stability, and release kinetics while employing strategies to overcome biological barriers and enhance delivery to target sites. Cancer is recognized as one of the

leading causes of mortality worldwide, highlighting the critical importance of developing therapeutic drugs for controlling and treating tumors and preventing their metastasis.[3]

Targeted drug design for cancer can enhance early diagnosis and treatment efficacy. Using effective and selective drugs allows for the identification and treatment of cancer at its early stages, potentially reducing disease-associated mortality rates.[4] These drugs also stimulate advanced research in the fields of biological sciences and pharmacology. Investigations into the design and development of anticancer therapeutics can lead to significant progress in this area.[5]

Artificial intelligence (AI) and machine learning (ML) hold considerable potential in drug development, providing significant enhancements across multiple research stages. These include the identification of novel targets, improved understanding of disease-target relationships, drug candidate selection, protein structure prediction, molecular compound design, advanced exploration of disease mechanisms, and research on prognostic and predictive biomarkers. AI-assisted drug design demonstrates remarkable capabilities in data analysis and drug discovery, showcasing substantial potential in healthcare. Integrating AI tools into cancer therapy programs has led to improved treatment efficacy. Significantly, AI algorithms can predict the effects of genetic mutations on chemotherapy or radiotherapy sensitivity. AI-driven drug identification, including simulated safety screening and efficacy prediction, offers a cost-effective approach.[6]

AI-based drug design can facilitate the development of novel and innovative therapeutic strategies for cancer. By employing advanced and targeted drugs, it becomes possible to provide novel, improved treatments for cancer patients. Innovations across various stages of drug research are essential to establish robust manufacturing processes that meet quality specifications, ensure continuous production, minimize process deviations, and comply with regulatory standards.[7] Such innovations in pharmaceutical production help reduce financial risks, facilitate supply chains, and enable effective production of high-quality pharmaceutical products to meet the increasing global demand for medications. AI can analyze massive datasets, including medical records, reports, and clinical experiences, to determine the optimal personalized treatment for each patient. Advances in natural language processing enable the development of algorithms capable of studying drug effects on the human body. AI can also manage potential drug-drug interactions in patients receiving multiple medications by analyzing data to predict and mitigate such interactions.[8]

Intelligent drug design, utilizing advanced computational approaches such as computer-aided drug design (CADD) and AI, plays a fundamental role in the development of effective anticancer therapeutics. Researchers applying techniques such as molecular modeling, molecular docking, receptor mapping, and quantitative structure-activity relationships (QSAR) can design novel targeted anticancer drugs that act on key disease molecules.[7] AI plays a crucial role in identifying drug targets, screening large databases for interactions, and modeling drug-target interactions.[9]

CADD, through computational simulations, represents a promising alternative to traditional methods, enabling the development of personalized treatments and drug repurposing while minimizing ethical and financial limitations. Additionally, computational methods enable precise prediction of drug targets and facilitate the development of novel anticancer therapeutics. The design of anticancer peptides demonstrates the potential of recombinant proteins in creating targeted cancer treatments, enhancing drug efficacy and delivery specificity.[10]

Designing therapeutic drugs for cancer is of critical importance because cancer is a complex and high-risk disease with significant societal and individual health impacts. Through development and research in cancer pharmacology, it is possible to provide more effective, efficient, and less toxic treatment options.[11] Due to the complexity and heterogeneity of cancer across different types and stages, drug design requires innovative and advanced approaches that enhance therapeutic outcomes and reduce cancer-related mortality.[12]

AI algorithms have significantly impacted anticancer drug design, providing faster and more effective solutions. These algorithms contribute to reducing research costs, accelerating drug development processes, and improving the accuracy and speed of cancer diagnosis and treatment decision-making.[13] By leveraging AI technologies such as machine learning (ML) and deep learning (DL), researchers can enhance current approaches to anticancer drug research, yielding improved health outcomes and more efficient cancer management. Integrating AI into drug design not only streamlines processes but also holds the promise of revolutionizing future cancer therapy strategies.[14]

Drug design is a lengthy and time-consuming process, encompassing multiple stages from target discovery to clinical trials. Across these stages, several computational methods can be employed to assist in drug discovery. By employing mathematical modeling and accurate predictions, AI can accomplish tasks in the biological sciences previously considered impossible. This system functions as a technology-based platform, utilizing a range of advanced tools and networks for predictive approaches. The fields of medicine and drug delivery device design are among the multiple domains benefiting from AI applications.[15] Figure 1 illustrates a comprehensive overview of computational techniques utilized at different stages of drug discovery.[16]

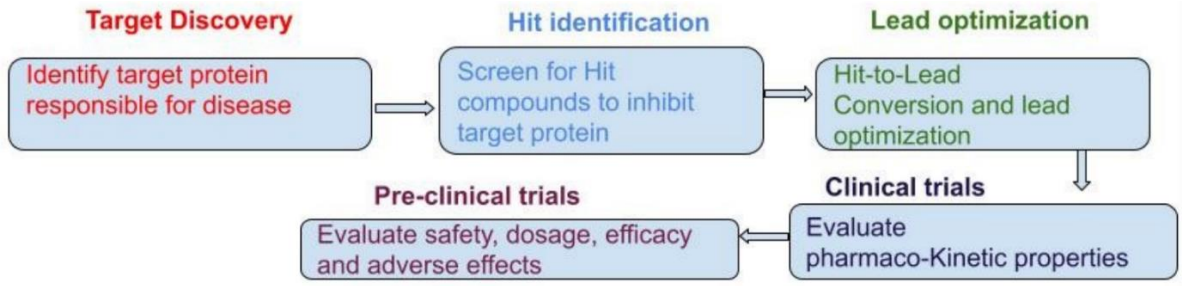


Figure 1. Flowchart of Drug Design Process.[16]

Designing therapeutic drugs for cancer can help prevent disease progression and metastasis, thereby reducing cancer-associated risks and contributing to disease prevention. Personalized medicine is expected to become more widespread in the future, although some clinicians may be hesitant to adopt rapid changes in existing clinical practices.[17] Given the critical importance of therapeutic drug design for cancer, research and development in this domain is of paramount significance. Such efforts can advance pharmaceutical and medical sciences, improve patients' quality of life, and enhance hope for effective cancer treatment.

2. Computer-Aided Drug Design

Computer-aided drug design (CADD) is a powerful tool that utilizes computational techniques for drug design, prediction of biological activity, and optimization of pharmacological properties.[18] By saving time and resources through approaches such as virtual screening, structure-based drug design, and ligand-based drug design, CADD has transformed the drug discovery process.[19] Figure 2 illustrates a schematic overview of some prominent approaches and platforms in computer-aided drug discovery and development. These approaches and platforms include target identification, docking-based virtual screening, combinatorial sampling, scoring functions, molecular similarity calculations, virtual library design, and sequence-based drug design. These elements are highly interconnected, where improvement in one aspect can enhance the performance of others.[16]

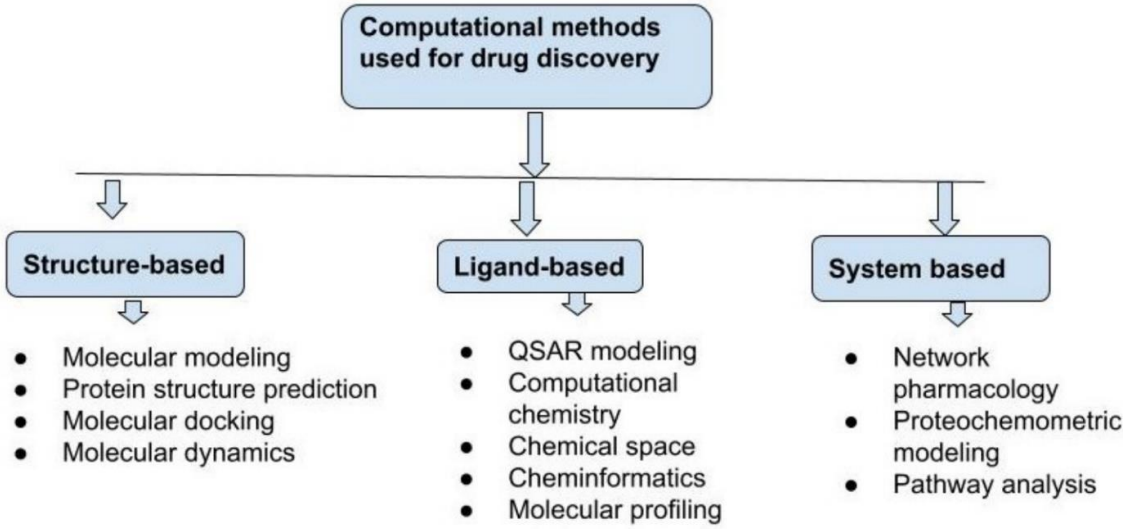


Figure 2. Categories of Different Computational Tools Used for Drug Discovery.[16]

2.1. Structure-Based Drug Discovery (SBDD)

Structure-based drug design is one of the oldest techniques in drug development. Recently, DeepMind released the AI system AlphaFold, capable of predicting the three-dimensional (3D) structure of a protein from its amino acid sequence. AlphaFold has predicted the 3D structures of all human proteins. The availability of a large number of target protein structures has significantly accelerated SBDD processes.[20] The massive volume of 3D structural data has driven data-driven SBDD. The Protein Data Bank (PDB) is the largest repository of biomolecular structural data globally, primarily derived from X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy techniques. In 1998, the number of structures registered in the PDB was 2,058.[21] Molecular docking and molecular dynamics have traditionally been employed as computational methods in SBDD for extended periods. Currently, advancements in

DL and cloud computing have provided new avenues for SBDD protocols.[22]

2.2. Ligand-Based Drug Discovery (LBDD)

Ligand-based drug design is employed when the three-dimensional structure of a target molecule is unknown. This approach is based on the principle that molecules bind to the desired biological target. LBDD operates according to the principle of "similarity-property relationship," which states that molecules with similar structures exhibit similar biochemical properties.[23] LBDD is a computational method used in drug development to identify potential drug candidates by screening large molecular databases for compounds with high similarity to known active molecules. It utilizes advanced computational algorithms to assess molecular shape and chemical properties, comparing them to known active compounds to accelerate the identification of promising drug candidates.[24] Additionally, LBDD plays a crucial role in predicting structure-activity relationships (SAR) of compounds and in designing novel drugs with enhanced bioactivity.[25]

LBDD using quantitative structure-activity relationship (QSAR) modeling involves several steps: (1) selection of chemical datasets; (2) generation of molecular descriptors; (3) construction of QSAR models by splitting datasets into training and test sets; (4) model validation and virtual ligand design; (5) prediction of optimal ligands and assessment of QSAR model accuracy; and (6) experimental validation of selected compounds.[26]

Three primary approaches are widely used to identify novel molecules in ligand-based drug design (Figure 3). The first method, termed ligand chemical similarity, selects new molecules with high structural, physicochemical, and affinity similarity to known active compounds. The second method, pharmacophore mapping, identifies functional groups that interact with targets and uses this information to design more potent molecules. The third method, QSAR modeling, correlates multiple features of chemical compounds with their biological activity toward the studied target.[27]

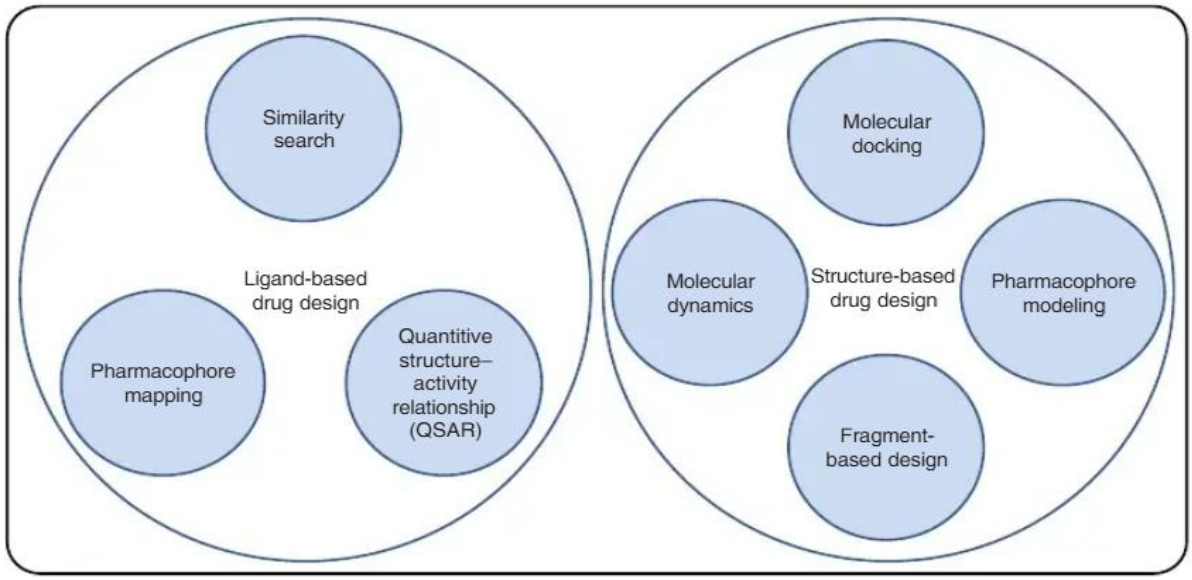


Figure 3. Main Methodologies for *in silico* Ligand- and Structure-Based Drug Design.[27]

2.3. System-Based Drug Discovery (SBDD)

The aim of system-based drug development is to conduct a comprehensive analysis of the genome, proteome, and their interactions to understand the positive or negative effects of chemical compounds on these functions.[28] In computational drug design, the conventional approach involves identifying a target protein and subsequently attempting to find a protein inhibitor. However, a protein functions as a chemical entity interacting with other proteins, forming a protein-protein interaction (PPI) network. When proteins interact, they are influenced by the chemical environment of the interacting proteins. In system-based drug design, researchers aim to inhibit an "interaction module" rather than a single protein. This module may constitute part of an interaction network or proteins within a specific metabolic pathway. Advances in multi-omics data analysis and network biology have led to the emergence of the field known as "network medicine." Biological networks are particularly important, as complex diseases are often the result of interactions among multiple proteins or genetic factors.[29]

This computational drug design approach relies on digital repositories for analyzing and modeling compounds, aiding in the identification of potential drug candidates and understanding SAR.[30] In disease-specific contexts such as cancer and schizophrenia, CADD has been fundamental in targeted therapies by identifying novel drug targets and facilitating the discovery of new drugs with improved efficacy and reduced toxicity.[31]

Computer-aided drug discovery methods have contributed

to the production of small-molecule therapeutics for over three decades. These techniques can be categorized as either structure-based or ligand-based. In theory, high-throughput screening (HTS) and structure-based approaches are similar. Ligand-based techniques predict activity based on the structural similarity or dissimilarity of candidate ligands to known active ligands.[32]

Traditional HTS assays often require extensive development and validation before implementation. Since CADD requires substantially less preparation time, researchers can simultaneously conduct in silico drug design studies while preparing HTS assays. The concurrent use of both methods provides a dual advantage for high-throughput drug discovery projects.[33] Currently, CADD plays a significant role in discovering pharmacologically active compounds that have successfully passed clinical trials and have been developed into novel therapeutics for various diseases.[34]

With advancements in computational technologies and AI, it has become feasible to design new and more efficient drugs for cancer treatment. Overall, CADD represents a powerful approach that enhances the drug discovery and development process, significantly supporting the pharmaceutical industry, although certain limitations still exist that require further development.[35] Table 1 describes some features, applications, performance, limitations, advantages, and disadvantages of CADD.

Table 1. Features, Applications, Effectiveness, Limitations, Advantages, and Disadvantages of CADD[36]

Section	Additional Information (Content)
Features	Use of computer simulations to predict and optimize the structure and function of drugs - Acceleration and cost reduction of the drug discovery and development process - Ability to discover new drugs with high accuracy and efficiency - Capability to overcome challenges associated with traditional drug design
Applications	Drug discovery for various diseases, including cancer, infections, and autoimmune diseases - Design of new drugs with improved properties, such as increased solubility or absorption - Optimization of existing drugs to reduce side effects or increase efficacy - Engineering of enzymes and metabolic pathways for various industrial applications
Effectiveness	CADD can significantly accelerate the drug discovery and development process and reduce its costs - CADD can lead to the discovery of new drugs with high accuracy and efficiency that would otherwise not have been possible - CADD can help overcome challenges related to traditional drug design, such as poor solubility or weak absorption of drugs
Limitations	CADD is a complex tool that requires deep expertise and knowledge in various fields - CADD is not always reliable or accurate and may predict incorrect results - Some potentially active compounds may not be identified by CADD
Advantages	Reduction in drug development time and cost - Increased likelihood of discovering new and effective drugs - Reduction in the risk of drug toxicity - Possibility of designing more precise drugs with fewer side effects
Disadvantages	Requires specialized expertise and equipment - May not always be reliable or accurate - Some potentially active compounds may be overlooked

3. CADD and AI Algorithms for Cancer Therapy

Technology-based AI systems can replicate human intelligence by utilizing a suite of advanced tools and networks. AI technologies are increasingly employed at various stages of the drug discovery process to save time and enhance profitability. These applications include real-time cell sorting, cell classification,[37] computation of molecular properties based on quantum mechanics (QM),[38] computational organic synthesis, generation of novel compounds, and a wide range of other tasks.[39,40]

AI, through enhanced automation, can process vast amounts of data, which encourages its widespread adoption.[41] Multiple ML algorithms, such as Support Vector Machines (SVM) and Random Forests (RF), have been applied in drug design to bridge the gap between random and rational drug discovery.[42]

AI provides excellent opportunities for target identification, peptide synthesis, toxicity and property evaluation, physicochemical studies of drugs, and monitoring drug efficacy and repurposing.[43] Some of the applications of AI in the pharmaceutical industry are illustrated in Figure 4.

In the context of designing drugs for cancer therapy, AI and DL algorithms play a crucial role. These AI methods can assist in analyzing complex data, identifying patterns and relationships within datasets, and predicting the efficacy of drugs in cancer treatment.[45] Below, some of the AI and DL algorithms used in cancer drug design are discussed.

3.1. ML Algorithms

ML refers to the process of training computers to understand data and make decisions based on patterns and examples. This process may involve supervised learning, unsupervised learning, and reinforcement learning.[46] These algorithms utilize large and complex datasets for prediction and information analysis. They can detect various patterns in the data and leverage them for the development of new drugs and enhancement of treatment processes.[47] In ML, any type of data can be used to develop predictive models. Model development typically involves data cleaning, outlier removal, and imputation of missing values, followed by feature selection, key variable identification, algorithm selection, model development, and model evaluation.[48]

3.1.1. Supervised ML

Supervised ML involves training algorithms to classify

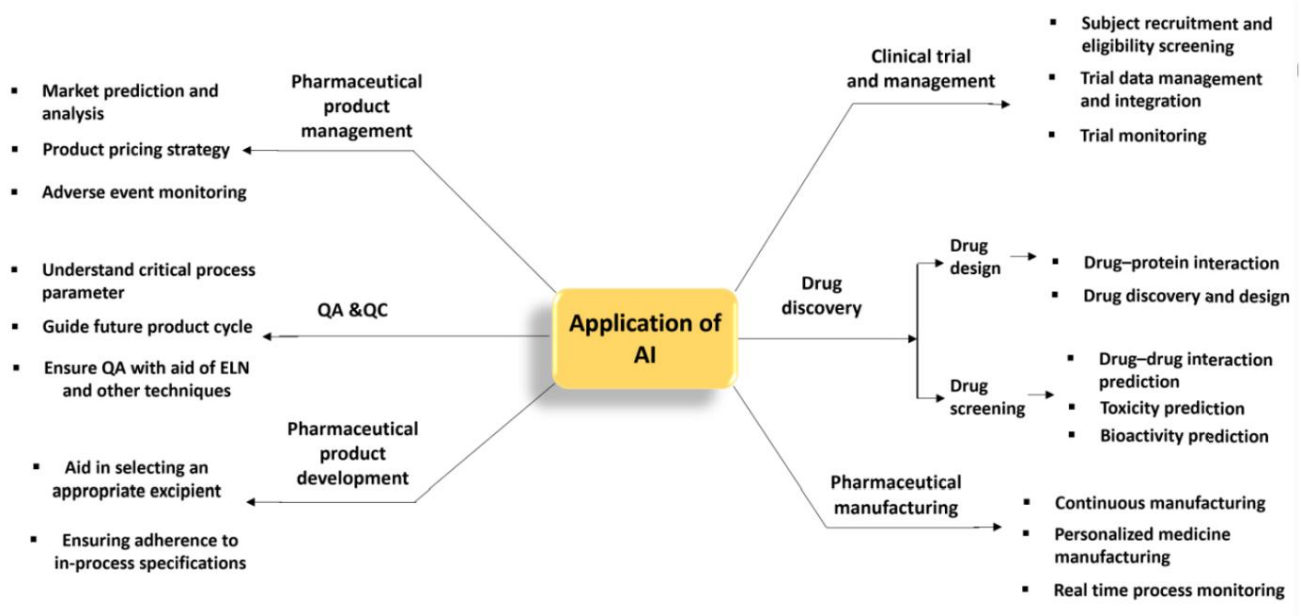


Figure 4. Application of AI in Multiple Sub-disciplines of the Pharmaceutical Industry, Including Pharmaceutical Product Management and Drug design.[44]

data or accurately predict outcomes using labeled datasets. Upon receiving input, the model updates its weights using reinforcement techniques to ensure proper alignment. Supervised learning is widely employed across industries and organizations for a variety of practical tasks. Its two primary types are classification and regression.[49]

3.1.2. Unsupervised ML

Unsupervised learning is applied when only input variables are available and no predefined outputs exist. The main objective is to understand data distributions to gain deeper insights. Unsupervised learning can be divided into two major categories: ensemble learning and clustering.[49]

3.1.3. Reinforcement Learning

Reinforcement learning (RL) is an intriguing subset of ML that has garnered significant attention in both scientific and commercial domains. Unlike supervised and unsupervised learning, RL is a continuous learning approach that maintains autonomy and can rapidly adapt to changing environments.[50] In some classification tasks, supervised learning may outperform RL; however, RL can continuously learn with minimal or no human intervention, making it a highly desirable approach.[51]

3.2. Deep Learning

Deep learning (DL), a subfield of ML, allows systems to learn independently from unlabeled and unstructured data. The term "deep learning" was coined in the late 20th century to describe artificial neural networks (ANNs) by Igor Aizenberg and colleagues.[52] DL utilizes ANNs with multiple processing layers.[46]

These algorithms leverage neural architectures to simulate human brain behavior. They can identify complex patterns within data and use them to predict drug efficacy. Neural network diagrams effectively capture interactions and structural information among atoms and molecules, which can be applied to predict protein–ligand binding affinity.[53]

Deep neural networks with attention mechanisms, such as Graph Attention Networks (GATs), transformers, and large language models (LLMs) including Bidirectional Encoder Representations from Transformers (BERT), pre-trained generative transformers (GPT),[54] and Bidirectional and Auto-Regressive Transformers (BART), have become pivotal tools in drug discovery and development.[55,56] By selectively focusing on specific segments of sequences, these models can extract salient features, providing valuable insights into structure–activity relationships critical for drug design.[57] Recent studies have demonstrated that this inherent capacity renders attention-based neural networks more interpretable than conventional DL techniques.[58] Moreover, attention mechanisms facilitate parallel processing, substantially improving computational efficiency.[59] These networks perform exceptionally in tasks such as generating novel molecular structures,[60] predicting drug–target interactions,[61] identifying drug-binding sites,[62] and predicting toxicity.[63] They also contribute significantly to protein structure prediction,[64] gene expression analysis,[65] sequence analysis,[66] genomic analysis,[67] proteomics,[68] spatial transcriptomics,[69] and pharmacotherapy,[70] ultimately advancing precision medicine and personalized treatments.

Generative adversarial networks (GANs) represent a novel ML approach consisting of two competing neural networks: a generator and a discriminator. The discriminator

aims to classify and distinguish real data from synthetic data produced by the generator. The generator produces synthetic data by leveraging feedback from the discriminator, trained on labeled real data. This optimization process iterates until a local Nash equilibrium is achieved, at which point further cost reduction in both networks ceases.[71]

In cheminformatics and CADD, innovative applications of GANs have emerged.[72] Variants such as Conditional GANs[73] and Wasserstein GANs[74] have been particularly effective for tasks including novel molecule generation[75] and optimization of molecules with desired properties.[76]

3.3. Big Data Analytics Algorithms

Big data analytics algorithms are employed to model and analyze extensive laboratory and clinical datasets, aiding researchers in better understanding cancer biology and designing effective therapeutics.[77] The integration of big data and AI promises a transformative impact on our understanding of cancer, from its onset to screening, diagnosis, treatment, response, toxicity, recurrence, and survival.[78]

AI is increasingly embedded in various aspects of cancer research, including standardization of large datasets and biobanks, elucidation of modifiable risk factors, identification of biomarkers or novel drug targets, development of predictive models and knowledge graphs, and establishment of new service platforms. These foundational components enable efficient collection and utilization of cancer-related big data. Nonetheless, numerous challenges remain in areas such as data harmonization, missing data management, and big data governance (Table 2).

Table 2. List of Challenges and Potential Solutions for Efficient Management and Use of Big Data.[79]

Main Category	Challenges	Solutions
Data Collection	Volume and Complexity: The massive volume and inherent complexity of data, including genetic, clinical, and lifestyle information, can make processing and analysis challenging.	Employ scalable computational infrastructure and advanced algorithms
	Data Quality Variability: Data quality varies from rigorous quality control in some datasets to potential inaccuracies in others.	Implement standardized data quality assessment procedures
	Quality and Standardization: Inconsistencies in data quality and lack of standardization due to varying definitions, different measurements, and temporal changes from evolving clinical guidelines and recording practices can lead to unreliable results.	Define clear variable definitions and measurements, implement dynamic algorithms to handle temporal changes, and adopt global data collection standards
	Integration: Integrating diverse data types from multiple sources (e.g., genomic data, imaging, electronic health records — EHR) is a complex task.	Develop integrated platforms and adopt interoperable/common data models
Data Management	Privacy Preservation: Ensuring the privacy and security of sensitive patient data is a major concern.	Implement strong encryption techniques and strict privacy policies
	Collaboration: Barriers to data sharing between institutions can hinder overall research progress.	Form data-sharing consortia and establish collaboration agreements
	Ethical Issues: Navigating the complex landscape of regulations and ethical considerations surrounding patient data use is challenging.	Provide clear guidelines, ethical oversight, and ongoing consultation with legal experts
Model Interpretation	Interpretability: The lack of comprehensive metadata and advanced analysis methods may result in uninterpretable models or lack of robustness.	Apply comprehensive metadata standards to improve data context, combine AI with domain expertise, and use interpretable models
	Clinical Application: AI interpretations must integrate seamlessly into existing clinical workflows.	Develop user-friendly interfaces and provide real-time decision support
	Algorithmic Bias: Algorithms can inherit biases from training data, potentially leading to unfair or incorrect outcomes.	Use diverse training datasets, perform regular evaluations, and adjust algorithms for fairness

CADD and AI are transforming the pharmaceutical industry by accelerating drug discovery and development processes. CADD approaches, including ML, Bayesian networks, and probabilistic reasoning, enhance traditional drug design methods by generating chemical libraries, optimizing candidate drugs, and evaluating their properties.[80] Both CADD and AI provide cost-effective and time-efficient solutions for designing anticancer drugs. However, AI is particularly promising for the faster, cheaper, and more efficient development of novel therapeutics.[81]

While CADD focuses on structure-based design and library screening, AI's data processing capabilities allow for the identification of new biological relationships, improved chemical design, increased success rates, and accelerated innovation testing.[82] Table 3 presents a comparison between CADD and AI-assisted drug design.

4. Computer-Aided Design of Recombinant Protein-Based Drugs for Cancer Therapy

Recombinant proteins are produced using genetic engineering technologies and can be specifically targeted to cancer cells. These proteins can activate various mechanisms to combat cancer cells, including tumor growth

Table 3. Comparison between CADD and AI

Criterion	CADD	AI Drug Design (AI-Based)
Approach	Computational approach: Uses computational methods and computer simulations for designing and developing new drugs.	Uses AI techniques such as ML, neural networks, and DL in drug design.
Data Reliance	Relies on existing laboratory and clinical data for modeling and prediction.	Relies on broader and more extensive data, including biochemical, pharmacological, and medical information.
Prediction Accuracy	Relatively high ability to predict drug properties, but still needs improvement.	Drug property prediction accuracy has dramatically improved.
Efficiency	The design process is relatively time-consuming and requires skilled experts.	The design process is faster and more efficient, requiring fewer experts.
Level of Automation	Very high level of automation and mechanization.	Relatively limited level of automation and mechanization.
Scalability	Limited scalability due to heavy computations and high computational resource requirements.	High scalability and ability to process massive volumes of data.
Handling Complexity	Limited ability to handle molecular and systemic complexities.	High ability to handle molecular and systemic complexities.

inhibition, induction of apoptosis, and enhancement of immune responses.[83] Recombinant protein-based therapeutics have been successfully applied in clinical settings.

CADD plays a crucial role in the development of novel drugs by utilizing computational tools and techniques. This includes predicting biological activity, pharmacological properties, and potential toxicity of drug candidates, thereby aiding the design of targeted drug delivery systems.[19] By incorporating recombinant proteins into this process, researchers can improve the accuracy of binding free energy predictions for ligands, resulting in the design of ligands with stronger and more selective binding.[84] Moreover, computational techniques such as molecular docking analyses can identify compounds that target specific proteins, potentially modulating their activity and serving as candidate therapeutic agents.[85] Overall, the integration of recombinant proteins in CADD holds significant promise for advancing drug discovery and development.[86]

Recombinant proteins play an essential role in computer-aided cancer therapy. Through advanced approaches like CADD, researchers can utilize databases containing key proteins and enzymes involved in cancer progression.[87] Integrating computational biomodeling has also led to the discovery of small-molecule inhibitors against therapeutic targets in cancers, underscoring the capability of recombinant protein-based drug design to develop more effective cancer therapies.[88]

The use of computational technologies and AI in designing recombinant protein-based drugs for cancer accelerates and optimizes the drug discovery and development process. This approach enables the creation of more targeted and effective therapeutics for cancer treatment. Table 4 summarizes recombinant protein-based drug design using computational techniques, highlighting key applications and advantages in this advanced and critical area of the pharmaceutical industry.

Table 4. Computational Techniques for the Design of Recombinant Protein Drugs[89]

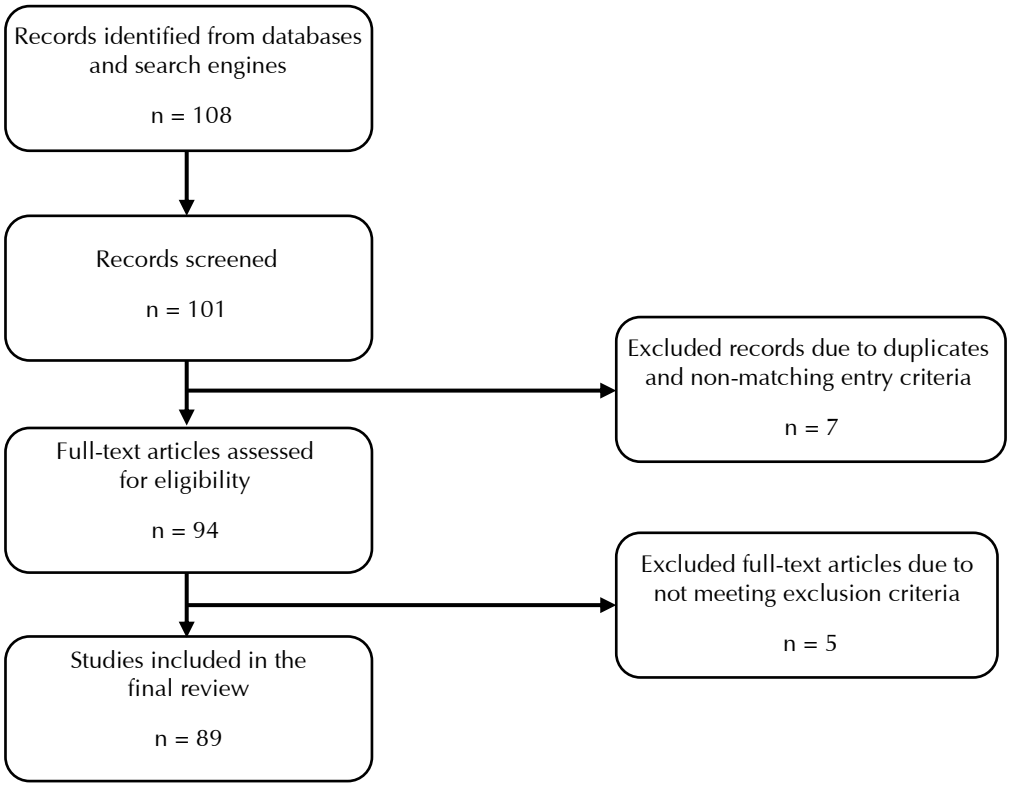
Topic	Description
Protein Engineering	Using computer modeling and simulation to design and engineer recombinant protein structures with optimized properties.
Binding Affinity Prediction	Employing molecular dynamics computations to accurately predict the binding strength of recombinant proteins to biological targets.
Stability Optimization	Using computer simulations to improve the physical and chemical stability of recombinant proteins.
Immunogenicity Prediction	Applying computational methods to predict the immunogenicity potential of recombinant proteins and design them to reduce it.
Rational Design	Using targeted computational approaches to design recombinant proteins with optimized efficacy and safety.
High-Throughput Screening	Utilizing cloud computing resources for fast and effective screening of large protein libraries.
Formulation Development	Using computer modeling to optimize the formulation of recombinant protein-based drugs.

5. Research Method (Systematic Review)

This study was conducted as a systematic review to examine the latest advancements and applications of CADD and AI algorithms in the development of protein-based anticancer drugs. Literature searches were performed in reputable databases such as PubMed, Scopus, ScienceDirect, Springer, Scholar, and SID using relevant keywords. For a more precise search, a list of

keywords used to define the role of AI algorithms in the healthcare domain was extracted, reviewed, and applied from published studies. Keywords included terms such as "computer-aided drug design," "artificial intelligence algorithms," "computer-aided cancer drug design," "recombinant proteins," and related combinations in both Persian and English. The search was limited to English-language articles published within the last ten years.

Article selection was based on reviewing titles and abstracts to evaluate their relevance to the study's focus on the use of CADD and AI algorithms in anticancer drug design. Duplicate articles and studies outside the scope of the topic were excluded. To extract and analyze data, key information from the articles, including methods, findings, results, and analyses of CADD methods and AI algorithms applied in the design of recombinant protein-based drugs for cancer treatment, was examined. Full texts of the selected articles were also reviewed, and relevant points were extracted. Additionally, the advantages, challenges, and advancements of each approach were assessed by evaluating and comparing the efficiency and impact of CADD and AI algorithms in accelerating drug development for cancer. Ultimately, the findings of this systematic review are presented in an analytical format.



Flowchart (Number of Retrieved Studies).

6. Results

Drug discovery and development is a multifaceted and complex endeavor. This intricate process comprises four core stages, beginning with target identification and validation, progressing through compound screening and lead optimization, and culminating in preclinical and clinical studies. The identification of pathophysiological factors and biological targets represents the first step in this process. Bioinformatics, genomics, and proteomics research are essential for discovering cellular and genetic targets. Initially, the first molecule exhibiting activity against the identified target is discovered. This can be achieved either through chemical libraries or by isolating natural products from bacteria, plants, and fungi.[90] Next, the lead molecule with the desired potential for drug development is identified. Lead optimization involves further modification of the selected lead molecule to enhance its specificity and efficacy, even at lower doses. The functional properties of newly synthesized therapeutic candidates are improved through iterative cycles combining cellular assays and structure–activity relationship (SAR) analyses. Subsequently, animal models are employed for in vivo studies, including pharmacokinetic analysis and toxicity assessment. Ultimately, following all preclinical evaluations, the drug candidate enters clinical trials, which determine whether the candidate provides the intended therapeutic benefits while ensuring patient safety.[91] This process is time-consuming and labor-intensive.

CADD enables the design of targeted drugs that interact with specific therapeutic targets (such as proteins involved in cancer pathogenesis) while optimizing pharmacological properties, including efficacy and safety. This approach accelerates the drug discovery process and reduces

development costs for anticancer therapeutics.[92]

Simultaneously, AI algorithms, which process complex biological data, play a significant role in anticancer drug design. AI, with its ability to detect hidden patterns and relationships in biological data, can predict drug effects on target proteins and optimize candidate compounds. These capabilities substantially enhance the efficiency of designing and developing recombinant protein-based anticancer drugs.[93]

Bibliometric analyses provide a comprehensive overview of publications in the field of CADD. Figure 5 presents a co-occurrence map of keywords for the years 2005–2024. During this period, 659 keywords were identified, and using VOSviewer software, the corresponding map was generated. The analysis revealed 42 clusters and 1,528 links. As shown in Figure 6, the thematic categories of AI, molecular docking, ML, and virtual screening exhibit the strongest connections with other topics and represent the most prominent research hotspots during the period under review.

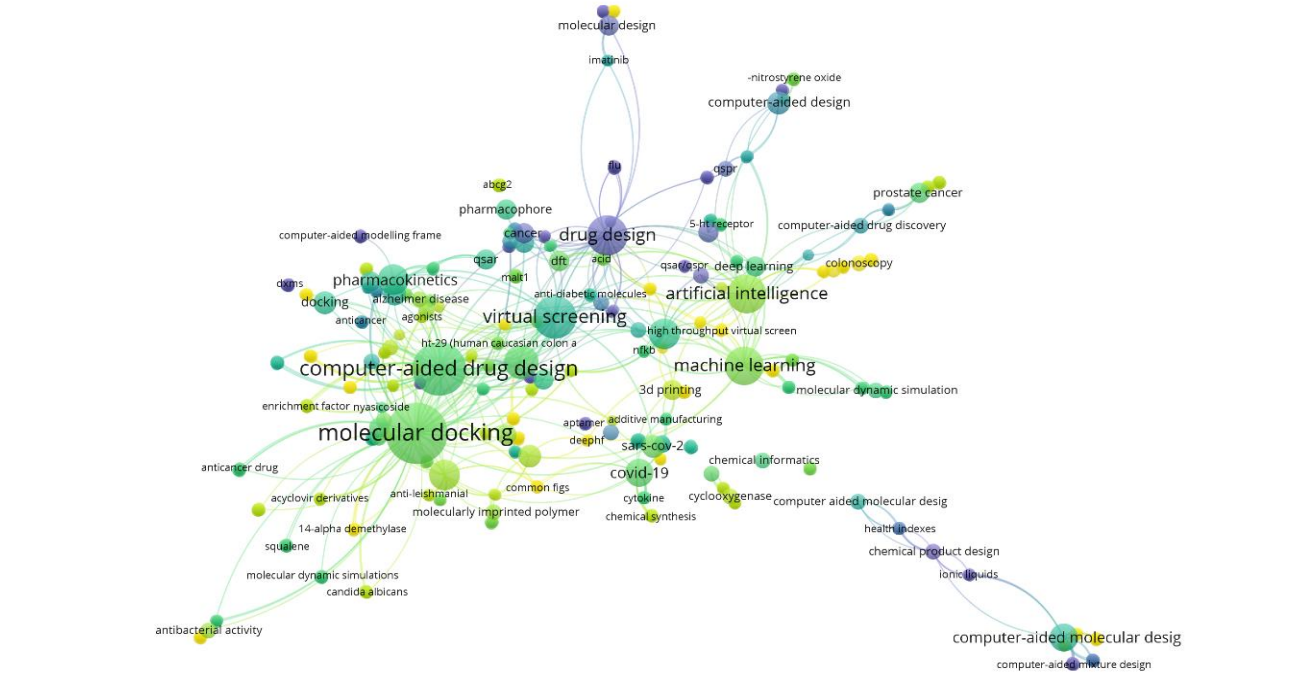


Figure 5. Co-occurrence Map of Keywords in the Field of CADD from 2005-2024.

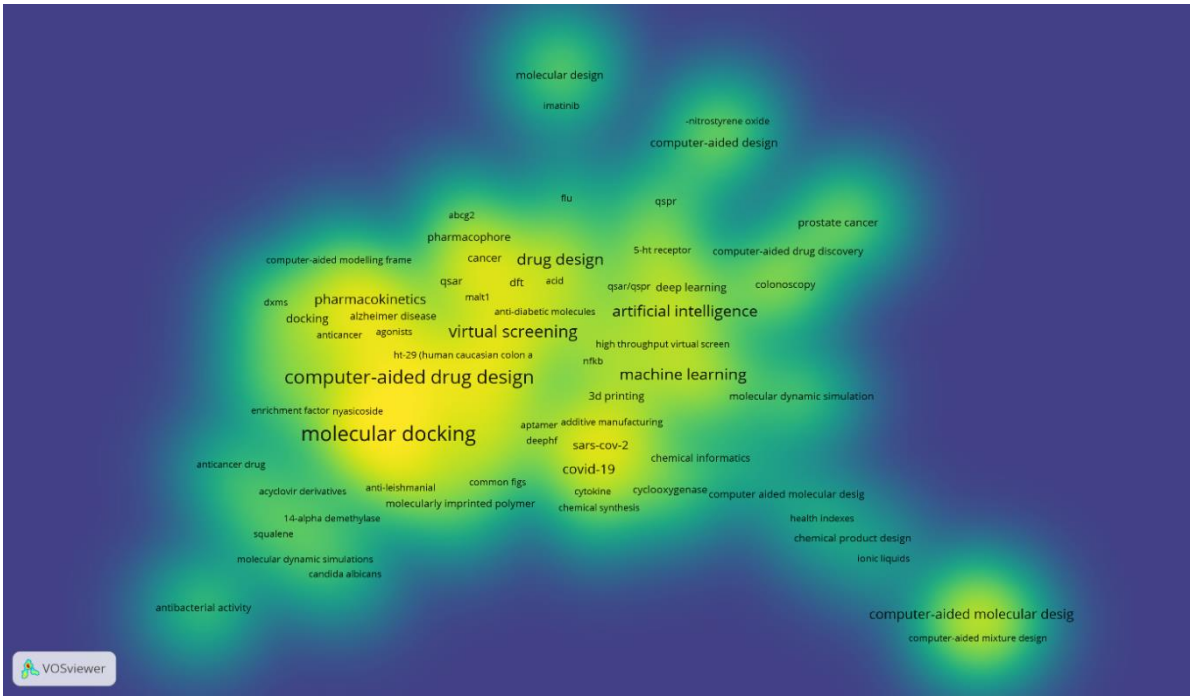


Figure 6. Co-occurrence Map of Keywords in the Field of CADD from 2005-2024.

6.1. Types of Computer-Aided Drug Design

The term "computer-aided drug design" (CADD) refers to molecular design using computational techniques to discover or develop pharmaceutical drugs. Recently, the application of computational chemistry and molecular modeling for CADD has expanded significantly across fields such as nanotechnology, molecular biology, and biochemistry. A key advantage of in silico drug design is that it renders drug research and development more cost-effective. This creative process focuses on discovering new drugs based on understanding a biological target—a strategy sometimes referred to as rational drug design.

A protein, for instance, represents a common biological molecule whose activity can be modulated by a drug to achieve therapeutic benefits for the patient. At its core, drug design involves creating compounds that interact with biomolecular targets in a shape- and charge-complementary manner.[94]

The main methods of CADD, as illustrated in Figure 7, can be categorized as follows:

Structure-Based Drug Design (SBDD): This approach encompasses molecular modeling and homology modeling, molecular docking, receptor-specific generation and mapping, molecular dynamics simulations, integrated methods, and in silico target prediction.

Ligand-Based Drug Design (LBDD): This includes quantitative structure–activity relationship (QSAR) studies, fragment-based descriptors, AI-assisted drug design, virtual screening, and HTS approaches—covering both virtual and experimental compound optimization.

Multi-Target Drug Design Strategies: In this approach, a biologically critical molecule involved in a specific metabolic or signaling pathway related to a disease or pathology, or contributing to pathogen virulence or survival, is recognized as a biomolecular target, often a protein or nucleic acid. Future drug targets should inherently possess the potential to treat or prevent diseases. Small molecules can be designed to enhance or inhibit target function in a disease-modifying pathway. Complementary compounds for the target binding site, such as receptor agonists, antagonists, inverse agonists, modulators, enzyme activators or inhibitors, or ion channel openers/blockers, can be developed accordingly.[95]

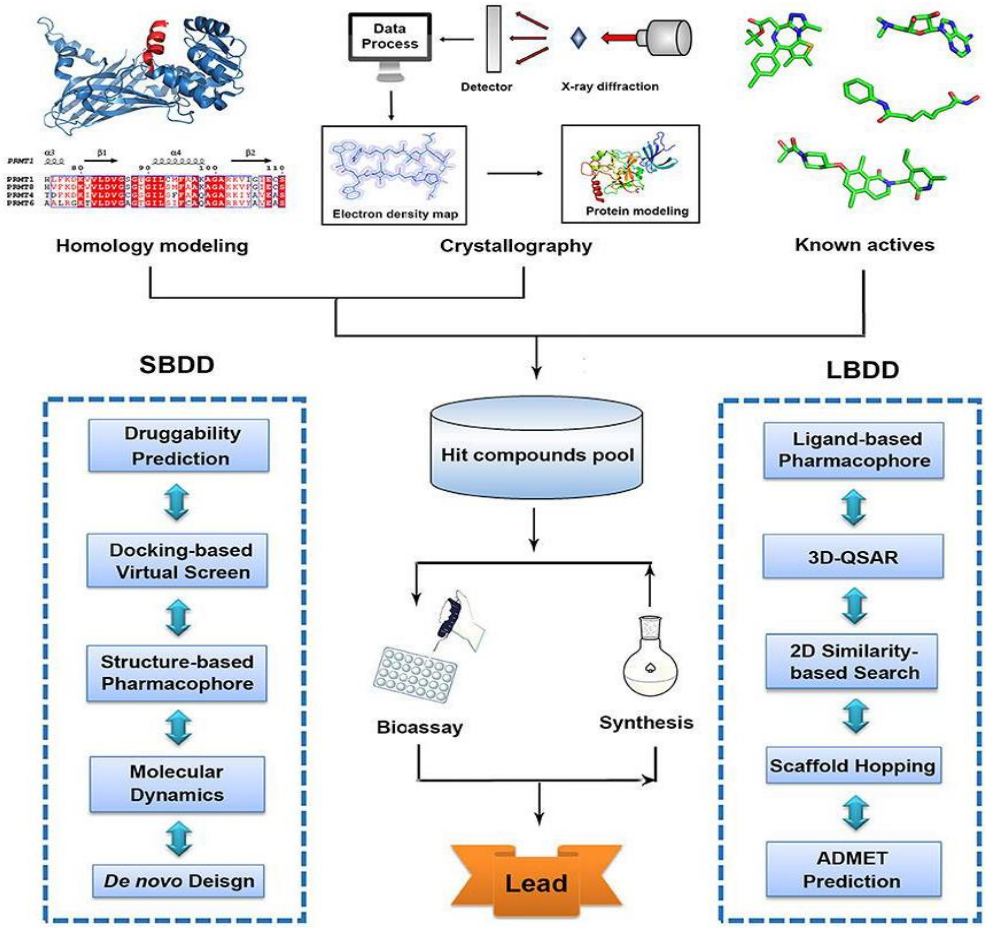


Figure 7. Traditional Workflow of SBDD and LBDD.[95]

Since drug–target interactions with off-target molecules may result in adverse side effects, small molecules (drugs) can be designed to avoid interactions with "off-target" molecules, often referred to as anti-targets.

Closely related targets identified via sequence homology exhibit the highest probability of cross-reactivity and, consequently, the greatest potential for side effects due to similarity at binding sites. Drugs are typically small organic molecules synthesized chemically. However, biologics, including polymer-based drugs, are increasingly produced through biological processes. Furthermore, therapeutic applications for mRNA-based gene-silencing technologies are also feasible.[95]

6.2. Big Data Sources in Drug Design

Large-scale data in drug discovery exists in diverse forms and types, which can be raw or processed, standardized or unstandardized. Extracting meaningful information from such heterogeneous data is challenging. Drug discovery relies on data from multiple domains, including clinical trials, bioassays, pharmacology, and structural biology. These datasets, generated from disparate fields and sources, constitute vast big data resources, where AI plays a crucial role in addressing the inherent complexity.[96] The continuous growth of big data necessitates more computational resources and advanced computational algorithms for analysis. The demand for higher computational power has shifted the practice from personal computers to high-performance computing systems, cloud computing, and graphics processing units (GPUs) to handle large-scale data analysis.[97]

The accumulated big data used in drug discovery can be organized into various categories or databases, including: chemical compound collections (e.g., PubChem, ChEMBL), drug-like compounds (e.g., DrugBank, e-Drug3D), drug targets, including genomic and proteomic data (e.g., BindingDB, Supertarget), bioassay, metabolism, and efficacy studies (e.g., HMDB, TTD), (Table 5).[98] Over the years, multiple data-sharing initiatives have emerged alongside the development of HTS technologies. Big data are required at various stages of the drug discovery process. The initial step involves screening large libraries of chemical compounds to identify potential lead drugs. The chemical compound library space is extensive, encompassing virtual, designed, and synthesized compounds, with properties and distributions described in both public and proprietary databases. Consequently, these data sources are massive, providing a wide range of multidimensional information for drug discovery and development, including chemical structures, bioassay data, target structures, and clinical data. The sheer volume and complexity of these datasets are increasing exponentially, creating opportunities for leveraging AI and ML to achieve rapid and effective solutions in drug discovery.

Table 5. Data Sources Used in Drug Discovery

No	Database		Data type	Ref
Chemical collections				
1	ChEMBL	https://www.ebi.ac.uk/chembl/	Chemical database containing bioactive and drug-like molecules	99
2	PubChem	https://pubchem.ncbi.nlm.nih.gov/	Chemical database containing chemicals and their activity against biological targets	100
3	Small molecule pathway database (SMPDB)	https://smpdb.ca/	Database containing small molecule pathway of humans	101
4	ZINC	http://zinc.docking.org/	Database containing curated chemical compounds	102
5	Human metabolome database (HMDB)	https://hmdb.ca/	Database containing 1) Chemical data, 2) clinical data, and 3) molecular biology/biochemistry data	103
6	Binding database (BindingDB)	https://www.bindingdb.org/bind/index.jsp	Database containing small molecules binding afnity data and protein targets	104
Drug/Drug-like compound				
1	DrugBank	https://go.drugbank.com	Database containing information about drug and drug targets	105
2	Drugs@FDA database	https://www.accessdata.fda.gov/scripts/cder/daf/	Database containing FDA approved drug molecules	106
3	Drug central	https://drugcentral.org/	Database containing active chemical entities, pharmaceutical product and drug mode of action	107
Drug Targets				
1	Supertarget	https://bioinformatics.charite.de/supertarget/	Database containing drug target information	108
2	Ligand Depot	http://ligand-depot.rutgers.edu/	Database containing information of chemical and structural information about small molecules	109
3	BioGRID	https://thebiogrid.org/	Database information about protein, genetic and chemical interactions	110
Protein–Protein Interaction				
1	Database of interacting proteins (DIP)	https://dip.doe-mbi.ucla.edu/dip/Main.cgi	Database containing information of protein–protein interaction	111

2	Therapeutic target database (TTD)	http://db.idrblab.net/ttd/	Database containing information of known therapeutic protein and nucleic acid targets	112
3	Potential drug target database (PDTD)	http://crdd.osdd.net/pmics.php	Database containing information about drug target with known 3d structure	113
Efficacy and assay screening				
1	BioCyc	https://biocyc.org/	Database containing information of organism specific Pathway/ Genome Database	114
2	BRENDA	https://www.brenda-enzymes.org/	Database containing information of enzyme function data	115
3	Reactome	https://reactome.org/	A curated database containing information of biological pathways, including the metabolic, protein trafficking and signalling pathways	116
4	KEGG	https://www.genome.jp/kegg/	Database containing information of genomic, chemical and functional information	117
Drug toxicity prediction				
1	Comparative toxicogenomics database	https://ctdbase.org/	Database containing information about environmental exposures to human health	118
2	TOXNE	https://www.nlm.nih.gov/toxnet/index.html	Database containing hazardous substance data	119
3	DrugMatrix	https://ntp.niehs.nih.gov/data/drugmatrix	Database of toxicogenomic reference resources	120

6.3. Computational AI Tools Available for Drug Design

The importance of computational software in drug design is evident from the early stages of drug discovery. Advances in software development and the widespread availability of these tools have created new opportunities to leverage them in research and learning processes. Open-source software has gained substantial popularity due to its accessibility and usability. Many researchers also share their programs on platforms such as GitHub to accelerate the drug discovery process and promote widespread adoption of these AI-based resources (Table 6). Currently, multiple DL frameworks are available to users, including TensorFlow, PyTorch, Keras, scikit-learn, MXNet, Gluon, Swift, and Chainer ONNX. These frameworks require high-performance computational resources across various platforms such as central processing units (CPU), graphics processing units (GPU), and tensor processing units (TPU).[121] The built-in libraries of these frameworks are based on DL and are applicable across diverse scientific and technological domains, including healthcare. Python-based libraries such as TensorFlow, PyTorch, Keras, and scikit-learn are extensively used in drug discovery, where massive datasets are common. TensorFlow, an open-source library developed by Google, enables the development of models for predicting molecular activity in compound datasets using DL approaches. Keras, built on top of TensorFlow, provides a user-friendly interface and simplified debugging capabilities, facilitating model development. PyTorch, another open-source DL library, is used to define and train models that capture complex drug–drug and drug–target interactions, accelerating the drug discovery process.[122] Scikit-learn, an open-source ML library, offers a user-friendly platform for classification, regression, and dimensionality reduction tasks. Other software, such as Weka, is also widely utilized for ML-based applications in drug discovery, including classification and clustering.

6.4. Prediction of Proteins with Specific Properties

Computational protein design has emerged in recent years as a field with the potential to significantly impact protein-based drug design. This field primarily involves redesigning regions of protein sequences to enhance the stability of a specified three-dimensional protein structure. However, it has expanded from core residue redesign to other regions of proteins, as well as the design of novel protein backbone structures. In recent years, proteins with novel binding functions, new enzymes, protein libraries, full-chain designs, and *de novo* protein folds have been among the most notable achievements. Nonetheless, challenges related to search and scoring remain unsolved, and many design processes need to be evaluated at much larger scales to assess their efficiency.[141]

Prediction of proteins with specific properties relies on carefully curated datasets, which also hold substantial value in practical applications. Proteins such as transcription factors or RNA-binding proteins, which function through binding to DNA or RNA, are of particular interest.[142] The identification of such proteins is crucial for understanding transcriptional and translational regulation. Consequently, studies have been dedicated to predicting DNA-binding proteins using ML approaches.[143] By employing position-specific scoring matrices (PSSM), hidden Markov models (HMMs), secondary structure information (DSSP), and structures predicted by AlphaFold2 to generate combined amino acid features, GraphSite has not only improved the

Table 6. AI Computational Tools for Drug Design

No	Tools		Url	Ref
1	AlphaFold	Predicts tertiary structure of a protein using deep neural network	https://deepmind.google/blog/alphafold	20
2	Chemputer	Give detailed recipe for compound synthesis	https://zenodo.org/records/1481731	123
3	Conv_qsar_fast	Predict molecular properties-based CNN method	https://github.com/connorcoley/conv_qsar_fast	124
4	Chemical VAE	Automated chemical design using variational autoencoder (VAE)	https://github.com/aspuru-guzik-group/chemical_vae	125
5	DeepChem	An open-source Python library uses a deep learning algorithm for compound identification	https://github.com/deepchem/deepchem	126
6	DeepTox	Predict the toxicity of chemical compounds using DL algorithm	http://www.bioinf.jku.at/research/DeepTox/	127
7	DeepNeuralNetQSAR	Predict molecular activity using multilevel deep neural network (DNN)	https://github.com/Merck/DeepNeuralNet-QSAR	128
8	DeltaVina	Predict small molecule binding afnity with drug with a combination of random forest (RF) and AutoDock scoring function	https://github.com/chengwang88/deltavina	129
9	Hit Dexter	Predict frequent hitter by using ML algorithm	http://hitdexter2.zbh.uni-hamburg.de	130
10	InnerOuterRNN	Predicts the physical, chemical and biological properties using inner- and outer recursive neural networks	https://github.com/Chemoinformatics/Inner OuterRNN	131
11	JunctionTree VAE	De novo molecule design using junction tree variational autoencoder (VAE)	https://github.com/wengong-jin/icml18-jtnn	132
12	Neural graph fngerprint	Predict the property of novel compounds using CNN	https://github.com/HIPS/neural-fingerprint	133
13	NNScore	Predict the afnity of protein–ligand interaction using neural network-based scoring function	http://www.nbcr.net/software/nnscore	134
14	ORGANIC	De novo design of organic molecule and polymer using ML algorithm	https://github.com/aspuru-guzik-group/ORGANIC	135
15	Open Drug Design Toolkit (ODDT's)	Cheminformatics' pipeline using random forest score (RF)-Score and NNScore	https://github.com/oddt/oddt	136
16	PotentialNet	Predict binding afnity using graph convolutional neural network (CNN)	https://pubs.acs.org/doi/full/10.1021/acscentsci.8b00507	137
17	PPB2	Predict the target of query molecule using nearest neighbour and ML algorithm	https://ppb2.gdb.tools/	138
18	QML	Python toolkit for quantum ML	https://www.qmlcode.org/	139
19	REINVENT	De novo design of molecule using recurrent neural network (RNN) and RL	https://github.com/MarcusOlivecrona/REINVENT	140

Table 7. Drug Target Protein Prediction Algorithms

No	Protein representation	Method	Accuracy	Ref
1	A descriptor encoding the structural and physicochemical properties of a protein	SVM	83.70%	145
2	Composition of the amino acid residues, Hydrophobicity, Polarity, polarizability, Charge, Solvent accessibility, Normalized van der Waals volume	SVM	84.00%	146
3	Three different sets of physicochemical properties	SVM	89.78%	147
4	Word2vec, auto covariance, Cojoint Triad	CNN and Traditional ML methods	89.55%	148
5	Amino acid composition, amphiphilic pseudo-amino acid composition, dipeptide composition, Composition-Transition-Distribution, pseudo amino acid composition	SVM	91.90%	149
6	Grouped amino acid composition (GDPC), reduced amino acid alphabet (RAAA), novel encoder pseudo amino acid segmentation (S-PseAAC)	ERT, XGB, RF	93.78%	150
7	Dictionary, dipeptide composition, tripeptide composition, Composition Transition-Distribution	CNN-RNN	92.40%	151
8	ESM1b, predicted contact map	GCN	95.00%	152

prediction of protein–nucleic acid binding, but also enabled the identification of binding sites.[144] In predicting proteins belonging to specific functional classes, the use of protein structure prediction tools, exemplified by AlphaFold2, offers significant advantages. Table 7 summarizes studies focused on predicting drug-target proteins and identifying target proteins based on ML approaches.

Sun et al. evaluated the performance of various ML algorithms for predicting druggable proteins, employing Word2Vec for protein sequence representation and demonstrating its potential in this field.[148] Chen et al.

utilized a pre-trained protein language model based on sequences (ESM1b) combined with a graph convolutional neural network classifier to develop an advanced sequence - based method for identifying drug-target proteins. Their comprehensive model, QuoteTarget, successfully identified 1,213 potential drug targets across the human proteome. Moreover, the authors employed Gradient-weighted Class Activation Mapping (Grad-CAM) to infer residue-level binding weights from trained networks.[152] Regarding classification algorithms, most drug-target protein prediction studies relied on traditional ML algorithms or shallow neural networks, achieving accuracies exceeding 90%. Notably, fewer studies utilized deep neural network architectures, likely due to the limited availability of extensive drug-target protein datasets for training.

The integration of CADD and AI into a unified workflow facilitates the identification and design of novel drugs with enhanced efficacy and safety. This convergence also contributes to reducing costs and accelerating the development of anticancer drugs, leading to increased success rates in various clinical evaluation stages.[153] Overall, the application of CADD and AI in designing recombinant protein-based anticancer drugs represents a significant advancement in cancer therapy.[154] In recent years, AI algorithms have played a critical role in improving and accelerating the design and development of anticancer therapeutics.[155] Multiple studies have demonstrated that AI can effectively assist in candidate drug identification, activity and toxicity prediction, and optimization of drug properties.[156]

7. Discussion

The incorporation of AI in drug development, from laboratory research to patient care, holds substantial potential. AI can support rational drug design, assist in determining optimal therapies for individual patients (personalized medicine), and enable efficient management and utilization of clinical data for future drug discovery. The chemical space, comprising over 10^60 atoms, offers virtually limitless opportunities for the development of novel drug molecules. However, a lack of fundamental innovation continues to make drug development time-consuming and expensive. AI provides a solution to address these challenges. Through AI, researchers can identify target and lead compounds, accelerate drug target validation, and simplify structure-based drug planning.[157] AI also reduces the resources required for HTS, decreasing the reliance on animal experiments.[53]

AI has played a pivotal role in advancing data-driven progress across all stages of drug design, facilitated by the development of ML. ML, with its extensive knowledge of chemical configurations, has evolved into DL, offering enhanced generalization capabilities and improved processing of big data. This advancement further strengthens the integration of AI innovations and computational technologies into drug development, facilitating research and discovery. Many diseases are strongly associated with protein structural abnormalities, and structure-based drug design strategies can identify small molecules capable of binding to functional sites within proteins.[55] However, determining three-dimensional (3D) protein structures is challenging and costly, necessitating computational prediction. Although a protein's one-dimensional (1D) amino acid sequence can determine much of its 3D structure, accurate predictions remain difficult due to the vast conformational search space.[56]

Technological advancements and reduced instrumentation costs have significantly increased data generation in both volume and diversity, facilitating the emergence of extensive data resources.[47] Big data, characterized by high volume and complexity, require collection, integration, and analysis to address complex biomedical and scientific challenges. This extensive data mining is crucial for revealing hidden patterns and is often referred to as big data analytics.[157]

AI, designed to mimic human cognitive functions, leverages these abundant data sources. AI has been applied to multiple challenges in the drug discovery pipeline, including designing and identifying novel drug molecules, drug repurposing, preclinical and clinical trials, and personalized medicine.[59] Big data can be processed in raw or preprocessed forms, standardized or non-standardized, and in various formats. Extracting meaningful information from this heterogeneous data is challenging. Drug discovery relies on data from diverse domains, including clinical, bioassay, pharmacological, and structural biology datasets, and AI plays a key role in addressing the inherent complexities of these heterogeneous datasets.[60] Big data can be analyzed using DL algorithms, in contrast to conventional ML approaches, and thus large, standardized datasets are essential for DL modeling.[61]

Advances in big data and AI have dramatically transformed approaches for accelerating drug development timelines. Implementing AI methods facilitates the systematic and cost-effective identification of drug candidates, significantly reducing development time. Computational capabilities and algorithms in drug discovery leverage existing data to enhance insights and analytical assessments, encompassing all stages from candidate identification to production processes within the pharmaceutical industry.[62]

Drug discovery remains a complex, lengthy, and multidisciplinary process. Before reaching the market, a drug must pass through multiple stages, each with specific requirements, costs, and timelines. Despite significant advances in understanding biological systems, identifying a novel therapeutic entity remains predominantly a time-consuming, costly, and complex task. AI represents a rapidly evolving field, incorporating areas

such as reasoning, knowledge representation, and ML. ML has been widely applied in drug discovery efforts involving large datasets, employing algorithms and methodologies to identify patterns and extract insights from complex data. AI algorithms can analyze vast datasets to identify potential drug targets and predict biological activity of compounds with unprecedented accuracy. In drug design and optimization, ML models facilitate the prediction of drug interactions and side effects, thereby accelerating the development of safer and more effective therapeutics. Advanced simulations and predictive models reduce dependency on experimental testing, optimizing the production pipeline. Clinical trial phases also benefit substantially from AI and ML, which assist in selecting suitable candidates based on genetic and clinical data, optimizing trial designs, and predicting patient responses, improving patient stratification.

8. Conclusion

The present review indicates that CADD and AI algorithms have significantly evolved in recent years, demonstrating substantial potential in enhancing drug discovery and development processes. The use of AI in CADD, particularly in cancer therapy, allows for faster and more precise identification of effective drug molecules and targeting of specific cancer-associated proteins. AI models, capable of analyzing vast datasets and predicting complex drug-protein interactions, not only reduce research and development costs and time, but also increase success rates in discovering new drugs.

Compared with traditional drug development processes, CADD holds both theoretical and clinical significance. Key points include:

8.1. Theoretical Significance

Speed and efficiency: CADD enables rapid screening of large compound libraries, reducing time and cost in drug development.

Precision and control: CADD provides greater control over molecular properties, facilitating the design of drugs with specific desirable features, crucial for targeting disease mechanisms and minimizing off-target effects.

Insights into molecular mechanisms: CADD offers precise insights into drug-target interactions at the molecular level, helping researchers understand and optimize mechanisms of action.

Reduction of animal testing: By predicting biological activity and toxicity *in silico*, CADD can reduce the need for in vivo experiments.

8.2. Clinical Significance

Identification of effective drugs: CADD assists in identifying drug candidates with higher probabilities of success, increasing the likelihood of developing effective therapies.

Minimizing adverse effects: By designing drugs with high target specificity, CADD reduces side effects and unwanted off-target interactions.

Personalized medicine: CADD enables the design of patient-specific therapies, considering individual genetic variations, facilitating precision medicine.

Improved patient outcomes: CADD supports the development of more effective drugs with fewer side effects, improving therapeutic outcomes.

Technological advances in AI and CADD have led to more effective drugs with reduced adverse effects. Nonetheless, challenges remain, including the need for high-quality data, interpretability of AI results, and extensive clinical validation. To fully leverage AI in drug design, further research and the development of innovative methodologies are required. This review provides a comprehensive overview of the current state and future perspectives of CADD and AI in cancer therapy, encouraging researchers to adopt these advanced techniques and emphasizing the importance of continued investigation in this domain.

9. Challenges and Future Prospects in AI-Driven Drug Discovery

Despite significant advancements, the application of AI in drug discovery faces challenges that must be addressed to fully realize its potential. A primary challenge is the quality and availability of data. Incomplete, heterogeneous, or low-quality datasets can compromise the accuracy of ML models. Additionally, the interpretability of complex AI models, particularly DL algorithms, remains challenging due to their opaque nature. Ethical concerns, including patient data privacy and informed consent, also present significant obstacles.

Nevertheless, ongoing technological advancements and the development of novel algorithms offer promising solutions to these challenges. AI is poised to play an increasingly central role in drug discovery, particularly through integration with emerging fields like computational biology and tissue engineering. This convergence could enable innovative therapeutic approaches for complex diseases, such as cancer and neurodegenerative disorders, fostering the development of targeted and personalized therapeutics.

AI and ML have significantly reduced drug development timelines and costs while improving the success rate of discovering novel therapies. By analyzing large-scale genomic, proteomic, and clinical datasets, these technologies provide actionable insights that enhance precision in decision-making. However, realizing AI's full potential requires interdisciplinary collaboration among computer scientists, biologists, and pharmaceutical experts. Such collaboration, leveraging diverse expertise, can drive innovation in drug discovery, leading to more effective and safer treatments.

Research Highlights

What Is Already Known?

- AI and ML have significantly advanced CADD by enabling the analysis of large-scale biological and chemical datasets.
- Techniques such as molecular docking, virtual screening, and QSAR modeling are established tools for identifying drug candidates and predicting drug-target interactions.
- Tools like AlphaFold2 have revolutionized three-dimensional (3D) protein structure prediction, facilitating the identification of drug-binding sites.
- HTS and computational protein design have improved the efficiency of drug discovery, particularly for complex diseases like cancer.

What Does This Study Add?

- This study provides a comprehensive review of AI and ML applications in CADD, highlighting their transformative impact on drug discovery processes, from preclinical to clinical stages.
- It elucidates how AI streamlines HTS, enhances 3D protein structure prediction, and improves drug-target identification, thereby reducing development timelines and costs.
- By integrating AI with emerging fields like computational biology and tissue engineering, the study highlights future prospects for developing innovative, targeted therapies, reshaping therapeutic paradigms for complex diseases.

Author Contributions

Authors contributed equally to this work.

Acknowledgments

The authors would like to thank the Clinical Research Development Unit of Baqiyatallah Hospital, for all their support and guidance during carrying out this study.

Conflict of Interest Disclosures

All authors declared that they have no conflict of interest.

References

1. Diamantis N, Banerji U. Antibody-drug conjugates—an emerging class of cancer treatment. Br J Cancer. 2016;114(2):362-7. doi:10.1038/bjc.2015.435
2. Wu M, Huang S. Magnetic nanoparticles in cancer diagnosis, drug delivery and treatment. Mol Clin Oncol. 2017;7(5):738-46. doi:10.3892/mco.2017.1399
3. Khodos I, Ray SD. Anticancer therapeutic agents. In: Wexler P, ed. Encyclopedia of Toxicology. 4th ed. Elsevier; 2024:549-65. doi:10.1016/B978-0-12-824315-2.01116-7
4. Noor U, Ahmed T, Khalid S, Khalid M, Sharif A, Noor I, et al. Nano-drug designing strategies and targeted applications for cancer treatment. Int J Pharm Integr Health Sci. 2023;4(1):68-87. doi:10.56536/ijpihs.v4i1.68
5. Gellerman G. Theranostic targeted drug delivery - new insights into diagnosis and treatment of cancer. Harefuah. 2022;161(1):42-8.
6. Dilnawaz F. Application of Artificial Intelligence for Designing Cancer Nanomedicine. In: Dilnawaz F, Behura AK, eds. Artificial Intelligence Based Cancer Nanomedicine: Diagnostics, Therapeutics and Bioethics. Bentham Science Publishers; 2022:4-15. doi:10.2174/9789815050561122010005
7. Masilamani K, Senthilnathan B, Manoyogambiga M, Gowri R, Vigneshwar M, Sathiyasundar R, et al. Techniques and tools for in silico drug design for the development of anticancer drugs. Int J Life Sci Pharma Res. 2023;13(5):P130-P148. doi:10.22376/ijlpr.2023.13.5.P130-P148
8. Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, Chavda VP. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. 2023;15(7):1916. doi:10.3390/pharmaceutics15071916
9. Wang L, Song Y, Wang H, Zhang X, Wang M, He J, et al. Advances of artificial intelligence in anti-cancer drug design: a review of the past decade. Pharmaceuticals. 2023;16(2):253. doi:10.3390/ph16020253
10. Niveditha VG, Sindhu V, Kizhakedathil MPJ, Sundari IS, Balasubramaniyan M. Tool and techniques on computer-aided drug design for targeted cancer therapy. In: Malviya R, Sundram S, eds. Targeted Cancer Therapy in Biomedical Engineering. Singapore: Springer; 2023:781-829. doi:10.1007/978-981-19-9786-0_23
11. Kiyonga EM, Kekani LN, Chidziwa TV, Kahwenga KD, Bronkhorst E, Milne M, et al. Nano- and crystal engineering approaches in the development of therapeutic agents for neoplastic diseases. Crystals. 2022;12(7):926. doi:10.3390/cryst12070926
12. Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. Cancer Discov. 2021;11(4):900-915. doi:10.1158/2159-8290.CD-21-0090
13. Berkat W. The Impact of Artificial Intelligence in Cancer Research and Treatment. OSF Preprints. 2023. doi:10.31219/osf.io/3fpw5
14. Ali M, Wani SUD, Khan AA, Dey T, Salahuddin M. Current status of artificial intelligence in cancer management: emerging trends, limitations and future outlook. Research Square. 2023. doi:10.21203/rs.3.rs-2724591/v1
15. Choudhuri S, Mallik S, Ghosh B, Si T, Bhadra T, Maulik U, et al. A Review of Computational Learning and IoT Applications to High-Throughput Array-Based Sequencing and Medical Imaging Data in Drug Discovery and Other Health Care Systems. In: De S, Das R, Bhattacharyya S, eds. Applied Smart Health Care Informatics: A Computational Intelligence Perspective. Hoboken, NJ: John Wiley & Sons. 2022:83-109. doi:10.1002/9781119743187.ch5
16. Choudhuri S, Yendluri M, Poddar S, Li A, Mallick K, Mallik S, et al. Recent Advancements in Computational Drug Design Algorithms through Machine Learning and Optimization. Kinases Phosphatases. 2023;1(2):117-40. doi:10.3390/kinasesphosphatases1020008
17. Das KP, J C. Nanoparticles and convergence of artificial intelligence for targeted drug delivery for cancer therapy: Current progress and challenges. Front Med Technol. 2023;4:1067144. doi:10.3389/fmedt.2022.1067144
18. Jambhulkar S. Review on Computer-Aided Drug Designing for Targeted Drug Delivery Systems. J Clin Med Res. 2023;5(3):76-82. doi:10.37191/Mapsci-2582-4333-5(3)-134
19. Lin Y. Review of Modern Computer-aided Drug Design Methods. Int J Biol Life Sci. 2022;1(1):47-50. doi:10.54097/ijbls. v1i1.3230
20. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. Nature.

2020;577(7792):706-10. doi:10.1038/s41586-019-1923-7
21. Chen S, Yan K, Liu B. PDB-BRE: A ligand-protein interaction binding residue extractor based on Protein Data Bank. Proteins. 2024;92(1):145-53. doi:10.1002/prot.26596
22. Muscat S, Errico S, Danani A, Chiti F, Grasso G. Leveraging Machine Learning-Guided Molecular Simulations Coupled with Experimental Data to Decipher Membrane Binding Mechanisms of Aminosterols. bioRxiv. 2024. doi:10.1101/2024.01.31.578042
23. Nguyen VTD, Nguyen ND, Hy TS. Complex-based Ligand-Binding Proteins Redesign by Equivariant Diffusion-based Generative Models. bioRxiv. 2024. doi:10.1101/2024.04.17.589997
24. Sharma A. Ligand Based Virtual Screening Testing and Analysis for Plasmodium Falciparum. International Journal of Scientific Research in Engineering and Management. 2023;7(10):1-11. doi:10.55041/IJSREM25981
25. Chandraghatgi R, Ji H-F, Rosen GL, Sokhansanj BA. Streamlining Computational Fragment-Based Drug Discovery through Evolutionary Optimization Informed by Ligand-Based Virtual Prescreening. J Chem Inf Model. 2024;64(9):3826-40. doi:10.1021/acs.jcim.4c00234
26. Acharya C, Coop A, Polli JE, MacKerell AD. Recent Advances in Ligand-Based Drug Design: Relevance and Utility of the Conformationally Sampled Pharmacophore Approach. Curr. Comput-Aided Drug Des. 2011;7(1):10-22. doi:10.2174/157340911793743547
27. Makrynitsa GI, Lykouras M, Spyroulias GA, Matsoukas MT. *In silico* Drug Design. In: eLS. Chichester: John Wiley & Sons, Ltd. 2018:1-7. doi:10.1002/9780470015902.a0028112
28. Pujol A, Mosca R, Farrйs J, Aloy P. Unveiling the role of network and systems biology in drug discovery. Trends Pharmacol Sci. 2010;31(3):115-23. doi:10.1016/j.tips.2009.11.006
29. Barabбsi A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011;12(1):56-68. doi:10.1038/nrg2918
30. Niazi SK, Mariam Z. Computer-Aided Drug Design and Drug Discovery: A Prospective Analysis. Pharmaceuticals. 2024;17(1):22. doi:10.3390/ph17010022
31. Wang G, Bai Y, Cui J, Zong Z, Gao Y, Zheng Z. Computer-Aided Drug Design Boosts RAS Inhibitor Discovery. Molecules. 2022;27(17):5710. doi:10.3390/molecules27175710
32. Geronikaki A, Dubey G, Petrou A, Kirubakaran S. Computer-aided drug design: An overview. In: Roy K, ed. Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development. London: Academic Press. 2023:39-68. doi:10.1016/B978-0-443-18638-7.00024-4
33. Saudar SS, Surana SS, Shaharwale SS. Computer aided drug designing in development of herbal drug. International Int J Pharmacogn Pharm. 2023;5(1A):41-7. doi:10.33545/27067009.2023.v5.i1a.109
34. Salman MM, Al-Obaidi Z, Kitchen P, Loreto A, Bill RM, Wade-Martins R. Advances in Applying Computer-Aided Drug Design for Neurodegenerative Diseases. Int J Mol Sci. 2021;22(9):4688. doi:10.3390/ijms22094688
35. Wang YQ, Hu KF, Hu CJ. Application and development of systems biology in computer-aided drug design. Zhongguo Zhong Yao Za Zhi. 2023;48(11):2868-75. doi:10.19540/j.cnki.cjcmm.20230227.601
36. del Carmen Quintal Bojyrquez N, Campos MR. Traditional and Novel Computer-Aided Drug Design (CADD) Approaches in the Anticancer Drug Discovery Process. Curr Cancer Drug Targets. 2023;23(5):333-45. doi:10.2174/1568009622666220705104249
37. Nitta N, Sugimura T, Isozaki A, Mikami H, Hiraki K, Sakuma S, et al. Intelligent Image-Activated Cell Sorting. Cell. 2018;175(1):266-276.e213. doi:10.1016/j.cell.2018.08.028
38. von Lilienfeld OA. Quantum Machine Learning in Chemical Compound Space. Angew Chem Int Ed Engl. 2018;57(16):4164-9. doi:10.1002/anie.201709686
39. Coley CW, Barzilay R, Jaakkola TS, Green WH, Jensen KF. Prediction of Organic Reaction Outcomes Using Machine Learning. ACS Cent Sci. 2017;3(5):434-43. doi:10.1021/acscentsci.7b00064
40. 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
41. Reddy D. Data Engineering Challenges in AI Automation. In: 2023 International Conference on Computing, Electronics & Communications Engineering (iCCECE). Piscataway, NJ: IEEE. 2023:107-112. doi:10.1109/iCCECE59400.2023.10238496
42. Lai S. The Applications of Machine Learning to Novel Drug Discovery. Highlights Sci Eng Technol. 2023;76:642-8. doi:10.54097/zz7kz64
43. Zhong F, Xing J, Li X, Liu X, Fu Z, Xiong Z, et al. Artificial intelligence in drug design. Sci China Life Sci. 2018;61(10):1191-1204. doi:10.1007/s11427-018-9342-2
44. Dhudum R, Ganeshpurkar A, Pawar A. Revolutionizing Drug Discovery: A Comprehensive Review of AI Applications. Drugs Drug Candidates. 2024;3(1):148-71. doi:10.3390/ddc3010009
45. Belsare MH, Burusic J. A Conceptual Framework for Impact of Artificial Intelligence and Machine Learning (AIML) in Drug Development Within Pharmaceutical Industry. In: Editors not specified in available sources. Book title: Likely part of a Springer series or proceedings, e.g., "Advances in Intelligent Systems and Computing" or similar based on ISBN pattern. Singapore: Springer. 2023:291-307. doi:10.1007/978-981-99-2909-2_17
46. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19(1):64. doi:10.1186/s12874-019-0681-4
47. Farooq M, Shahid MU. Explainable AI and machine learning algorithms to predict treatment failures for patients with cancer. J Clin Oncol. 2023;41(16_suppl):e13577. doi:10.1200/JCO.2023.41.16_suppl.e13577
48. Carvalho TFM, Silva JCF, Calil IP, Fontes EPB, Cerqueira FR. Rama: a machine learning approach for ribosomal protein prediction in plants. Sci Rep. 2017;7(1):16273. doi:10.1038/s41598-017-16322-4
49. Manne R. Machine learning techniques in drug discovery and development. Int J Appl Res. 2021;7(4):21-8. doi:10.22271/allresearch.2021.v7.i4a.8455
50. Polydoros AS, Nalpantidis L. Survey of Model-Based Reinforcement Learning: Applications on Robotics. J Intell Robot Syst. 2017;86(2):153-73. doi:10.1007/s10846-017-0468-y
51. Quah KH, Quek C, Leedham G. Reinforcement learning combined with a fuzzy adaptive learning control network (FALCON-R) for pattern classification. Pattern Recognit. 2005;38(4):513-26. doi:10.1016/j.patcog.2004.08.011
52. Kubat M. Induction in multi-label domains. In: Kubat M, ed. An Introduction to Machine Learning. Cham:

Springer International Publishing; 2017:251-71. doi:10.1007/978-3-319-63913-0_13
53. Mathivanan JS, Dhayabaran VV, David MR, Karuna Nidhi M, Karuppasamy MP, Thenamudhan SS. Application of Deep Learning Neural Networks in Computer-Aided Drug Discovery: A Review. Curr Bioinform. 2024;19(9):851-8. doi:10.2174/0115748936276510231123121404
54. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT. Diagn Interv Imaging. 2023;104(6):269-74. doi:10.1016/j.diii.2023.02.003
55. Zhang S, Fan R, Liu Y, Chen S, Liu Q, Zeng W. Applications of transformer-based language models in bioinformatics: a survey. Bioinform Adv. 2023;3(1):vbad001. doi:10.1093/bioadv/vbad001
56. Liu Z, Roberts RA, Lal-Nag M, Chen X, Huang R, Tong W. AI-based language models powering drug discovery and development. Drug Discov Today. 2021;26(11):2593-607. doi:10.1016/j.drudis.2021.06.009
57. Cheng Z, Yan C, Wu FX, Wang J. Drug-Target Interaction Prediction Using Multi-Head Self-Attention and Graph Attention Network. IEEE/ACM Trans Comput Biol Bioinform. 2022;19(4):2208-18. doi:10.1109/TCBB.2021.3077905
58. Ren Q, Li M, Li H, Shen Y. A novel deep learning prediction model for concrete dam displacements using interpretable mixed attention mechanism. Adv Eng Inform. 2021;50:101407. doi:10.1016/j.aei.2021.101407
59. Lin T, Wang Y, Liu X, Qiu X. A survey of transformers. AI Open. 2022;3:111-132. doi:10.1016/j.aiopen.2022.10.001
60. Dollar O, Joshi N, Beck DA, Pfaendtner J. Attention-based generative models for de novo molecular design. Chem Sci. 2021;12(24):8362-72. doi:10.1039/D1SC01050F
61. Kalakoti Y, Yadav S, Sundar D. TransDTI: Transformer-Based Language Models for Estimating DTIs and Building a Drug Recommendation Workflow. ACS Omega. 2022;7(3):2706-17. doi:10.1021/acsomega.1c05203
62. Lee I, Nam H. Sequence-based prediction of protein binding regions and drug–target interactions. J Cheminform. 2022;14(1):5. doi:10.1186/s13321-022-00584-w
63. Teng S, Yin C, Wang Y, Chen X, Yan Z, Cui L, et al. MolFPG: Multi-level fingerprint-based Graph Transformer for accurate and robust drug toxicity prediction. Comput Biol Med. 2023; 164:106904. doi:10.1016/j.compbiomed.2023.106904
64. Zhou Y, Tan K, Shen X, He Z, Zheng H. A Protein Structure Prediction Approach Leveraging Transformer and CNN Integration. In: 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE). Piscataway, NJ: IEEE; 2024. doi:10.1109/ICAACE61206.2024.10548253
65. Yang F, Wang W, Wang F, Fang Y, Tang D, Huang J, et al. scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data. Nat Mach Intell. 2022;4(10):852-66. doi:10.1038/s42256-022-00534-z
66. Charoenkwan P, Nantasenamat C, Hasan MM, Manavalan B, Shoombuatong W. BERT4Bitter: a bidirectional encoder representation from transformers (BERT)-based model for improving the prediction of bitter peptides. Bioinformatics. 2021;37(17):2556-62. doi:10.1093/bioinformatics/btab133
67. Raad J, Bugnon LA, Milone DH, Stegmayer G. miRe2e: a full end-to-end deep model based on transformers for prediction of pre-miRNAs. Bioinformatics. 2022;38(5):1191-7. doi:10.1093/bioinformatics/btab823
68. Castro E, Godavarthi A, Rubinfien J, Givechian K, Bhaskar D, Krishnaswamy S. Transformer-based protein generation with regularized latent space optimization. Nat Mach Intell. 2022;4(10):840-51. doi:10.1038/s42256-022-00532-1
69. Zeng Y, Wei Z, Yu W, Yin R, Yuan Y, Li B, et al. Spatial transcriptomics prediction from histology jointly through Transformer and graph neural networks. Brief Bioinform. 2022;23(5):bbac297. doi:10.1093/bib/bbac297
70. Wang X, Xu X, Tong W, Roberts R, Liu Z. InferBERT: A Transformer-Based Causal Inference Framework for Enhancing Pharmacovigilance. Front Artif Intell. 2021;4:659622. doi:10.3389/frai.2021.659622
71. Fedus W, Rosca M, Lakshminarayanan B, Dai AM, Mohamed S, Goodfellow I. Many paths to equilibrium: GANs do not need to decrease a divergence at every step. arXiv preprint arXiv:1710.08446. 2017. doi:10.48550/arXiv.1710.08446
72. Lin E, Lin CH, Lane HY. Relevant applications of generative adversarial networks in drug design and discovery: molecular de novo design, dimensionality reduction, and de novo peptide and protein design. Molecules. 2020;25(14):3250. doi:10.3390/molecules25143250
73. Ge Q, Huang X, Fang S, Guo S, Liu Y, Lin W, et al. Conditional Generative Adversarial Networks for Individualized Treatment Effect Estimation and Treatment Selection. Front Genet. 2020;11:585804. doi:10.3389/fgene.2020.585804
74. Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. 2014. doi:10.48550/arXiv.1411.1784
75. Kadurin A, Nikolenko S, Khrabrov K, Aliper A, Zhavoronkov A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties *in Silico*. Mol Pharm. 2017;14(9):3098-104. doi:10.1021/acs.molpharmaceut.7b00346
76. Maziarka Ł, Pocha A, Kaczmarczyk J, Rataj K, Danel T, Warchoł M. Mol-CycleGAN: a generative model for molecular optimization. J Cheminform. 2020;12(1):2. doi:10.1186/s13321-019-0404-1
77. Kiryu Y. Medical Big Data Analysis Using Machine Learning Algorithms in the Field of Clinical Pharmacy. Yakugaku Zasshi. 2022;142(4):319-26. doi:10.1248/yakushi.21-00178-1
78. Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell. 2023;186(8):1772-91. doi:10.1016/j.cell.2023.01.035
79. Wu X, Li W, Tu H. Big data and artificial intelligence in cancer research. Trends Cancer. 2024;10(2):147-60. doi:10.1016/j.trecan.2023.10.006
80. Gini G. The impact of artificial intelligence methods on drug design. In: Roy K, ed. Cheminformatics, QSAR and Machine Learning Applications for Novel Drug Development. London: Academic Press; 2023:89-137. doi:10.1016/B978-0-443-18638-7.00009-8
81. Sahu A, Mishra J, Kushwaha N. Artificial Intelligence (AI) in Drugs and Pharmaceuticals. Comb Chem High Throughput Screen. 2022;25(11):1818-37. doi:10.2174/1386207325666211207153943
82. Singh R, Maddalwar S, Mishra M. Artificial Intelligence for Computer-Aided Drug Discovery. Drug Res. 2023;73(7):369-77. doi:10.1055/a-2076-3359
83. Ashwin A. An Investigation into the World of Protein-Based Therapeutics-Therapeutic Proteins. J Clin Med

Res. 2023;5(3). doi:10.37191/Mapsci-2582-4333-5(3)-135
84. Fajer M, Borrelli K, Abel R, Wang L. Quantitatively Accounting for Protein Reorganization in Computer-Aided Drug Design. J Chem Theory Comput. 2023;19(11):3080-90. doi:10.1021/acs.jctc.3c00009
85. Hameed I, Zubair A, Nazir A, Shahid K, Aimen S, Khan MI, et al. Computer-aided drug design against schizophrenia by targeting SP4. Biomed Lett. 2023;9(1):40-7. doi:10.47262/bl/9.1.20230501
86. Ece A. Computer-aided drug design. BMC Chem. 2023;17(1):26. doi:10.1186/s13065-023-00939-w
87. Ejiofor II, Ekeomodi CC, IlecChukwu AU, Ochiamu MC. CADD for Cancer Therapy: Current and Future Perspective. In: Malviya R, Sundram S, eds. Targeted Cancer Therapy in Biomedical Engineering. Singapore: Springer Nature Singapore; 2023:325-63. doi:10.1007/978-981-19-9786-0_9
88. Poonan P, Agoni C, Ibrahim MAA, Soliman MES. Glioma-Targeted Therapeutics: Computer-Aided Drug Design Prospective. Protein J. 2021;40(5):601-55. doi:10.1007/s10930-021-10021-w
89. Twite AA, Barnebey A, Brier LW, Barnett B, Jackson WM. Abstract 1525: A novel platform for protein engineering and modification to expand the therapeutic index of anti-tumor therapeutics. Cancer Res. 2023;83(7_Supplement):1525. doi:10.1158/1538-7445.Am2023-1525
90. Zhu T, Cao S, Su PC, Patel R, Shah D, Chokshi HB, et al. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. J Med Chem. 2013;56(17):6560-72. doi:10.1021/jm301916b
91. Chan HCS, Shan H, Dahoun T, Vogel H, Yuan S. Advancing Drug Discovery via Artificial Intelligence. Trends Pharmacol Sci. 2019;40(8):592-604. doi:10.1016/j.tips.2019.06.004
92. Bassani D, Moro S. Past, present, and future perspectives on computer-aided drug design methodologies. Molecules. 2023;28(9):3906. doi:10.3390/molecules28093906
93. Qian H, Wang B. Exploring the Mechanism of Inula japonica Thunb. against Non-small Cell Lung Cancer using a computer-aided drug design approach. Research Square. 2023. doi:10.21203/rs.3.rs-2970066/v1
94. Shaker B, Ahmad S, Lee J, Jung C, Na D. *In silico* methods and tools for drug discovery. Comput Biol Med. 2021;137:104851. doi:10.1016/j.compbiomed.2021.104851
95. Lu W, Zhang R, Jiang H, Zhang H, Luo C. Computer-Aided Drug Design in Epigenetics. Front Chem. 2018;6:57. doi:10.3389/fchem.2018.00057
96. de Souza A, Bittker JA, Lahr DL, Brudz S, Chatwin S, Oprea TI, et al. An Overview of the Challenges in Designing, Integrating, and Delivering BARD: A Public Chemical-Biology Resource and Query Portal for Multiple Organizations, Locations, and Disciplines. J Biomol Screen. 2014;19(5):614-27. doi:10.1177/1087057113517139
97. Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of Big Data challenges and analytical methods. J Bus Res. 2017;70:263-86. doi:10.1016/j.jbusres.2016.08.001
98. Pereira DA, Williams JA. Origin and evolution of high throughput screening. Br J Pharmacol. 2007;152(1):53-61. doi:10.1038/sj.bjp.0707373
99. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 2019;47(D1):D930-D940. doi:10.1093/nar/gky1075
100. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res. 2021;49(D1):D1388-D1395. doi:10.1093/nar/gkaa971
101. Frolkis A, Knox C, Lim E, Jewison T, Law V, Hau DD, et al. SMPDB: the Small Molecule Pathway Database. Nucleic Acids Res. 2010;38(suppl_1):D480-D487. doi:10.1093/nar/gkp1002
102. Irwin JJ, Shoichet BK. ZINC− a free database of commercially available compounds for virtual screening. J Chem Inf Model. 2005;45(1):177-82. doi:10.1021/ci049714+
103. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic Acids Res. 2007;35(suppl_1):D521-D526. doi:10.1093/nar/gkl923
104. Gilson MK, Liu T, Baitaluk M, Nicola G, Hwang L, Chong J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. Nucleic Acids Res. 2016;44(D1):D1045-D1053. doi:10.1093/nar/gkv1072
105. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P, et al. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34(suppl_1):D668-D672. doi:10.1093/nar/gkj067
106. Sahdeo S, Tomilov A, Komachi K, Iwahashi C, Datta S, Hughes O, et al. High-throughput screening of FDA-approved drugs using oxygen biosensor plates reveals secondary mitofunctional effects. Mitochondrion. 2014;17:116-25. doi:10.1016/j.mito.2014.07.002
107. Ursu O, Holmes J, Knockel J, Bologa CG, Yang JJ, Mathias SL, Nelson SJ, Oprea TI. DrugCentral: online drug compendium. Nucleic Acids Res. 2017;45(D1):D932-D939. doi:10.1093/nar/gkw993
108. Hecker N, Ahmed J, von Eichborn J, Dunkel M, Macha K, Eckert A, et al. SuperTarget goes quantitative: update on drug–target interactions. Nucleic Acids Res. 2012;40(D1): D1113-D1117. doi:10.1093/nar/gkr912
109. Feng Z, Chen L, Maddula H, Akcan O, Oughtred R, Berman HM, et al. Ligand Depot: a data warehouse for ligands bound to macromolecules. Bioinformatics. 2004;20(13):2153-5. doi:10.1093/bioinformatics/bth214
110. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34(suppl_1):D535-D539. doi:10.1093/nar/gkj109
111. Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the Database of Interacting Proteins. Nucleic Acids Res. 2000;28(1):289-91. doi:10.1093/nar/28.1.289
112. Chen X, Ji ZL, Chen YZ. TTD: Therapeutic Target Database. Nucleic Acids Res. 2002;30(1):412-5. doi:10.1093/nar/30.1.412
113. Gao Z, Li H, Zhang H, Liu X, Kang L, Luo X, et al. PDTD: a web-accessible protein database for drug target identification. BMC Bioinformatics. 2008;9(1):104. doi:10.1186/1471-2105-9-104
114. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, et al. BioCyc: a genomic and metabolic web portal with multiple omics analytical tools. FASEB J. 2019;33(S1):473.2. doi:10.1096/fasebj.2019.33.1_supplement.473.2
115. Schomburg I, Chang A, Schomburg D. BRENDA, enzyme data and metabolic information. Nucleic Acids Res. 2002;30(1):47-9. doi:10.1093/nar/30.1.47
116. Croft D, O'Kelly G, Wu G, Haw R, Gillespie M, Matthews L, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011;39(suppl_1): D691-D697. doi:10.1093/nar/gkq1018

117. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000;28(1):27-30. doi:10.1093/nar/28.1.27
118. Mattingly CJ, Rosenstein MC, Davis AP, Colby GT, Forrest JN Jr, Boyer JL. The Comparative Toxicogenomics Database: a cross-species resource for building chemical-gene interaction networks. Toxicol Sci. 2006;92(2):587-95. doi:10.1093/toxsci/kfl008
119. Fowler S, Schnall JG. TOXNET: information on toxicology and environmental health. Am J Nurs. 2014;114(2):61-3. doi:10.1097/01.NAJ.0000443783.75162.79
120. Ganter B, Snyder RD, Halbert DN, Lee MD. Toxicogenomics in drug discovery and development: mechanistic analysis of compound/class-dependent effects using the DrugMatrix® database. Pharmacogenomics. 2006;7(7):1025-44. doi:10.2217/14622416.7.7.1025
121. Nguyen G, Dlugolinsky S, Bobák M, Tran V, López García Б, Heredia I, et al. Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. Artif Intell Rev. 2019;52(1):77-124. doi:10.1007/s10462-018-09679-z
122. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: An imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 2019;32.
123. Steiner S, Wolf J, Glatzel S, Andreou A, Granda JM, Keenan G, et al. Organic synthesis in a modular robotic system driven by a chemical programming language. Science. 2019;363(6423):eaav2211. doi:10.1126/science. aav2211
124. Coley CW, Barzilay R, Green WH, Jaakkola TS, Jensen KF. Convolutional embedding of attributed molecular graphs for physical property prediction. J Chem Inf Model. 2017;57(8):1757-72. doi:10.1021/acs.jcim.6b00601
125. G ó mez-Bombarelli R, Wei JN, Duvenaud D, Hernández-Lobato JM, Sánchez-Lengeling B, Sheberla D, et al. Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent Sci. 2018;4(2):268-76. doi:10.1021/acscentsci.7b00572
126. Zhu H. Big data and artificial intelligence modeling for drug discovery. Annu Rev Pharmacol Toxicol. 2020; 60:573-89. doi:10.1146/annurev-pharmtox-010919-023324
127. Mayr A, Klambauer G, Unterthiner T, Hochreiter S. DeepTox: toxicity prediction using deep learning. Front Environ Sci. 2016;3:80. doi:10.3389/fenvs.2015.00080
128. Ghasemi F, Mehridehnavi A, Fassihi A, Pérez-Sánchez H. Deep neural network in QSAR studies using deep belief network. Appl Soft Comput. 2018; 62:251-8. doi:10.1016/j.asoc.2017.09.040
129. Wang C, Zhang Y. Improving scoring-docking-screening powers of protein–ligand scoring functions using random forest. J Comput Chem. 2017;38(3):169-77. doi:10.1002/jcc.24667
130. Stork C, Chen Y, Šícho M, Kirchmair J. Hit Dexter 2.0: machine-learning models for the prediction of frequent hitters. J Chem Inf Model. 2019;59(3):1030-43. doi:10.1021/acs.jcim.8b00677
131. Urban G, Subrahmanya N, Baldi P. Inner and outer recursive neural networks for chemoinformatics applications. J Chem Inf Model. 2018;58(2):207-11. doi:10.1021/acs.jcim.7b00384
132. Jin W, Barzilay R, Jaakkola T. Junction tree variational autoencoder for molecular graph generation. Proc Mach Learn Res. 2018;80:2323-32.
133. Duvenaud DK, Maclaurin D, Iparraguirre J, Bombarell R, Hirzel T, Aspuru-Guzik A, et al. Convolutional networks on graphs for learning molecular fingerprints. Adv Neural Inf Process Syst. 2015;28.
134. 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/ci200388 9
135. Sanchez-Lengeling B, Outeiral C, Guimarres GL, Aspuru-Guzik A. Optimizing distributions over molecular space. An objective-reinforced generative adversarial network for inverse-design chemistry (ORGANIC). ChemRxiv. 2017. doi:10.26434/chemrxiv. 5309668.v3
136. Wyjcikowski M, Zielenkiewicz P, Siedlecki P. Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. J Cheminform. 2015;7(1):26. doi:10.1186/s13321-015-0078-2
137. Feinberg EN, Sur D, Wu Z, Husic BE, Mai H, Li Y, Sun S, Yang J, Ramsundar B, Pande VS. PotentialNet for molecular property prediction. ACS Cent Sci. 2018;4(11):1520-30. doi:10.1021/acscentsci.8b00507
138. Awale M, Reymond JL. Polypharmacology Browser PPB2: target prediction combining nearest neighbors with machine learning. J Chem Inf Model. 2019;59(1):10-17. doi:10.1021/acs.jcim.8b00524
139. Cho A. No room for error. Science. 2020;369(6500):130-3. doi:10.1126/science.369.6500.130
140. Blaschke T, Ar ú s-Pous J, Chen H, Margreitter C, Tyrchan C, Engkvist O, et al. REINVENT 2.0: an AI tool for de novo drug design. J Chem Inf Model. 2020;60(12):5918-22. doi:10.1021/acs.jcim.0c00915
141. Rosenberg M, Goldblum A. Computational protein design: a novel path to future protein drugs. Curr Pharm Des. 2006;12(31):3973-97. doi:10.2174/138161206778743655
142. Hudson WH, Ortlund EA. The structure, function and evolution of proteins that bind DNA and RNA. Nat Rev Mol Cell Biol. 2014;15(11):749-60. doi:10.1038/nrm3884
143. Shadab S, Alam Khan MT, Neezi NA, Adilina S, Shatabda S. DeepDBP: deep neural networks for identification of DNA-binding proteins. Inform Med Unlocked. 2020;19:100318. doi:10.1016/j.imu.2020.100318
144. Shi W, Singha M, Pu L, Srivastava G, Ramanujam J, Brylinski M. Graphsite: ligand binding site classification with deep graph learning. Biomolecules. 2022;12(8):1053. doi:10.3390/biom12081053
145. Han LY, Zheng CJ, Xie B, Jia J, Ma XH, Zhu F, et al. Support vector machines approach for predicting druggable proteins: recent progress in its exploration and investigation of its usefulness. Drug Discov Today. 2007;12(7):304-13. doi:10.1016/j.drudis.2007.02.015
146. Li Q, Lai L. Prediction of potential drug targets based on simple sequence properties. BMC Bioinformatics. 2007;8(1):353. doi:10.1186/1471-2105-8-353
147. Jamali AA, Ferdousi R, Razzaghi S, Li J, Safdari R, Ebrahimie E. DrugMiner: comparative analysis of machine learning algorithms for prediction of potential druggable proteins. Drug Discov Today. 2016;21(5):718-24. doi:10.1016/j.drudis.2016.01.007
148. Sun T, Lai L, Pei J. Analysis of protein features and machine learning algorithms for prediction of druggable proteins. Quant Biol. 2018;6(4):334-43. doi:10.1007/s40484-018-0157-2
149. Charoenkwan P, Schaduangrat N, Lio' P, Moni MA, Shoombuatong W, Manavalan B. Computational prediction and interpretation of druggable proteins using a stacked ensemble-learning framework. iScience. 2022;25(9):104883. doi:10.1016/j.isci.2022.104883
150. Wang Z, Combs SA, Brand R, Calvo MR, Xu P, Price

G, et al. LM-GVP: an extensible sequence and structure informed deep learning framework for protein property prediction. Sci Rep. 2022;12(1):6832. doi:10.1038/s41598-022-10775-y
151. Yu L, Xue L, Liu F, Li Y, Jing R, Luo J. The applications of deep learning algorithms on in silico druggable proteins identification. J Adv Res. 2022;41:219-31. doi:10.1016/j.jare.2022.01.009
152. Chen J, Gu Z, Xu Y, Deng M, Lai L, Pei J. QuoteTarget: a sequence-based transformer protein language model to identify potentially druggable protein targets. Protein Sci. 2023;32(2):e4555. doi:10.1002/pro.4555
153. Saifi I, Bhat BA, Hamdani SS, Bhat UY, Lobato-Tapia CA, Mir MA, et al. Artificial intelligence and cheminformatics tools: a contribution to the drug development and chemical science. J Biomol Struct Dyn. 2024;42(12):6523-41. doi:10.1080/07391102.2023.2234039
154. Brown N, Ertl P, Lewis R, Luksch T, Reker D, Schneider N. Artificial intelligence in chemistry and drug design. J Comput Aided Mol Des. 2020;34(7):709-15. doi:10.1007/s10822-020-00317-x
155. Shirazi A, Tofighi M, Gharavi A, Gomez G. The application of artificial intelligence to cancer research: a comprehensive guide. Technol Cancer Res Treat. 2024;23:15330338241250324. doi:10.1177/15330338241250324
156. Tran TTV, Surya Wibowo A, Tayara H, Chong KT. Artificial intelligence in drug toxicity prediction: recent advances, challenges, and future perspectives. J Chem Inf Model. 2023;63(9):2628-43. doi:10.1021/acs.jcim.3c00200
157. Silva JCF, Teixeira RM, Silva FF, Brommonschenkel SH, Fontes EPB. Machine learning approaches and their current application in plant molecular biology: a systematic review. Plant Sci. 2019;284:37-47. doi:10.1016/j.plantsci.2019.03.020