Chinese Journal of Natural Medicines

Artificial intelligence in natural products research

Xiao Yuan, Xiaobo Yang, Qiyuan Pan, Cheng Luo, Xin Luan, Hao Zhang

Citation: Xiao Yuan, Xiaobo Yang, Qiyuan Pan, Cheng Luo, Xin Luan, Hao Zhang, Artificial intelligence in natural products research, Chinese Journal of Natural Medicines, 2025, 23(11), 1342−1357. doi: 10.1016/S1875−5364(25)60902−2.

View online: https://doi.org/10.1016/S1875−5364(25)60902−2

Chinese Journal of Natural Medicines 23 (2025) 1342-1357

Contents lists available at ScienceDirect

Chinese Journal of Natural Medicines

journal homepage: www.cjnmcpu.com/

Review

Artificial intelligence in natural products research

Xiao Yuan[a,Δ], Xiaobo Yang[b,Δ], Qiyuan Pan[a], Cheng Luo[b,c,d], Xin Luan[a,*], Hao Zhang[a,*]

[a] *Shanghai Frontiers Science Center of TCM Chemical Biology, Institute of Interdisciplinary Integrative Medicine Research, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China*

[b] *Zhongshan Institute for Drug Discovery, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Zhongshan 528437, China*

[c] *State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China*

[d] *State Key Laboratory of Functions and Applications of Medicinal Plants, Guizhou Medical University, Guiyang 550014, China*

ARTICLE INFO

Article history:
Received 12 February 2025
Revised 13 May 2025
Accepted 29 May 2025
Available online 20 November 2025

Keywords:
Natural products
Artificial intelligence
Deep learning
Drug discovery
Model interpretability

ABSTRACT

Artificial intelligence (AI) has emerged as a transformative technology in accelerating drug discovery and development within natural medicines research. Natural medicines, characterized by their complex chemical compositions and multifaceted pharmacological mechanisms, demonstrate widespread application in treating diverse diseases. However, research and development face significant challenges, including component complexity, extraction difficulties, and efficacy validation. AI technology, particularly through deep learning (DL) and machine learning (ML) approaches, enables efficient analysis of extensive datasets, facilitating drug screening, component analysis, and pharmacological mechanism elucidation. The implementation of AI technology demonstrates considerable potential in virtual screening, compound optimization, and synthetic pathway design, thereby enhancing natural medicines' bioavailability and safety profiles. Nevertheless, current applications encounter limitations regarding data quality, model interpretability, and ethical considerations. As AI technologies continue to evolve, natural medicines research and development will achieve greater efficiency and precision, advancing both personalized medicine and contemporary drug development approaches.

1. Introduction

Natural products (NPs) are bioactive compounds extracted from nature, found abundantly in plants, microorganisms, marine organisms, and other living entities [1](Fig. 1). Throughout human history, NPs have been fundamental to civilization's development. Their distinctive chemical structures and diverse functional properties render them essential in pharmaceutical research and development, healthcare, agriculture, environmental protection, and the food industry [2-4].

In healthcare, natural plants, herbs, and other naturally sourced substances serve multiple therapeutic purposes, including immune system enhancement, anti-aging effects, and sleep quality improvement. Many countries legally classify herbal medicines and plant-based drugs as dietary supplements [5]. Research demonstrates that fish oil supplementation reduces C-reactive protein levels in blood dialysis patients, thereby enhancing immunity [6]. Regarding sustainable agriculture and environmental protection, NPs such as biopesticides and plant extracts serve essential functions. Plant and microorganism-derived biopesticides, including azadirachtin and matrine, demonstrate high effectiveness, low toxicity, and biodegradability, thus reducing environmental pollution and pesticide residues in agricultural products [7]. In food industry applications, NPs provide various solutions for processing, preservation, and healthy diet promotion. Plant polyphenols, for instance, extend food shelf life while effectively reducing oil oxidation and preventing spoilage [8]. The pharmaceutical field notably benefits from NPs, as evidenced by landmark drug discoveries such as penicillin, paclitaxel, and avermectin [9]. Since 1981, approximately 35% of FDA-approved small-molecule drugs have originated from NPs [10].

Scientific and technological advancements have transformed NP research from traditional empirical methods and basic extraction techniques to more systematic and refined approaches [11, 12]. Despite extensive history in natural drug development and ongoing exploration of pharmacological mechanisms through modern scientific methods, significant challenges persist. The chemical complexity and diversity of natural medicines, often exhibiting synergistic effects, make isolating single effective components extremely challenging, resulting in extended research cycles and high costs [13]. NP extraction and separation processes demand substantial time, resources, and labor. Seasonal variations and environmental factors introduce process instability. Additionally, resource depletion and biodiversity threats impede research progress [14, 15]. Many natural medicines' pharmacological mechanisms remain incompletely understood, particularly at the molecular level. Their multi-component nature produces multi-target ef-

* Corresponding author.
E-mail addresses: luanxin@shutcm.edu.cn (X. Luan); zhanghao@shutcm.edu.cn (H. Zhang)

Δ These authors contributed equally to this work.

https://doi.org/10.1016/S1875-5364(25)60902-2

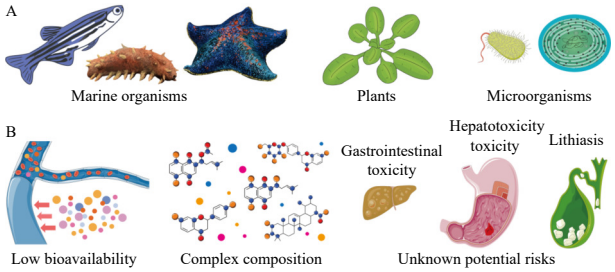

Fig. 1 Overview of challenges in natural product-based drug research. (A) Natural products are derived from a wide variety of sources, including plants, microorganisms, marine organisms, and other living systems. (B) Challenges in natural product-based drug research include low bioavailability, complex chemical compositions, and unclear safety or toxicity profiles.

fects, complicating research and clinical applications. These factors present both scientific and practical standardization challenges [16, 17].

Safety and bioavailability represent crucial concerns in natural medicine research. While generally considered milder, natural medicines can produce varying individual responses, and drug interactions may present risks [18, 19]. Limited bioavailability often characterizes natural medicines, and improving absorption and efficacy remains challenging [20, 21]. Comprehensive natural medicine development requires addressing pharmacological mechanism complexities, component analysis, and implementing advanced technologies to enhance safety, stability, and bioavailability (Fig. 1). Scientists continue exploring novel research methods and technological approaches to advance natural medicine development.

Artificial intelligence (AI) encompasses computer systems designed to perform tasks requiring human intelligence, including learning, reasoning, and pattern recognition [22]. AI development spans three historical stages. The first stage, symbolic AI (1950s to 1980s), featured rule-based expert systems and logic programming. The second stage (1980s to 2010s) marked the emergence of statistical methods and machine learning (ML), including support vector machines (SVM), random forests (RF), and Bayesian methods. The third stage (2010s to present) represents the deep Learning Revolution, characterized by neural network expansion through big data and Graphics Processing Unit (GPU) acceleration [22, 23]. While symbolic AI primarily represents historical development, it maintains relevance in drug discovery through knowledge graphs, particularly in connecting natural compounds to molecular targets and disease phenotypes [24, 25]. AI applications in NP research are extensive, enabling rapid classification of plant-derived molecules and protein targets in virtual screening and docking. AI facilitates automated mass spectrometry (MS)/nuclear magnetic resonance (NMR) spectra matching with known compounds to minimize rediscovery. Graph neural networks (GNNs) analyze enzyme steps in microbial or plant secondary metabolism for biosynthetic pathway prediction. Furthermore, AI guides optimization of cultivation conditions and gene editing to enhance desired NP yields [26].

The advanced capabilities of AI in virtual screening [27] and biological activity prediction [28] have substantially reduced development timelines and associated costs. The ongoing expansion of open databases and large-scale omics resources enables AI models to synthesize dispersed information through multimodal learning and knowledge graph frameworks [29], facilitating systematic and interpretable exploration of NP space. While conventional computational and experimental approaches, such as single target screening or isolated synthesis activities, enhance efficiency in specific steps, they remain less effective than AI's data-driven reasoning in overall innovation and understanding of polypharmacology [30-33]. Consequently, AI has emerged as a crucial technology for addressing persistent challenges and identifying novel natural medicines.

Recent years have witnessed significant progress in AI technology, bringing renewed momentum to NPs research. Specifically, advances in ML, deep learning (DL, and data analysis have enabled AI to excel in processing big data, recognizing complex patterns, and developing predictive models (Fig. 2). These AI-driven approaches have demonstrated remarkable potential in NP drug screening, pharmacological mechanism elucidation, bioactive component analysis, and synthetic route optimization, leading to notable improvements in research efficiency and cost reduction [34, 35]. Despite challenges such as data quality inconsistencies and model interpretability issues [36, 37], ongoing technological progress indicates AI's increasingly vital role in NP medicine development (Fig. 3).

This review presents a comprehensive analysis of AI applications in NP research. Through detailed examination of the convergence between AI and NP-based drug research, it aims to provide valuable insights for researchers and practitioners in related fields, supporting the efficient development and application of NP medicines. Table 1 presents several notable and successful applications of AI in NP drug discovery.

2. NP data: public repositories, challenges, and solutions

2.1. Public repositories for NP research

Public repositories of NPs serve as essential resources supporting research and applications related to natural compounds. Chemical Entities of Biological Interest (ChEBI), a prominent free database, focuses on small chemical compounds, cataloging distinct molecular entities including atoms, molecules, and ions. These entities comprise NPs and synthetic compounds utilized in biological processes, serving as a vital resource for chemistry, biology, and pharmacology research [38]. PubChem, maintained by the National Institutes of Health (NIH), functions as an open-access chemical database offering comprehensive information about small and large molecules, including their properties and bioactivity, serving scientists, students, and the general public [39].

Kyoto Encyclopedia of Genes and Genomes (KEGG) combines genomic, metabolic, and disease-related information, utilizing metabolic pathways and molecular networks to elucidate NPs' roles in biological systems, thus supporting NP research and drug development [40]. Natural Products Atlas specializes in microbial NPs, providing compound structures, biological classifications, synthesis, and ontological information, while offering chemical structure and text searches, and chemical diversity visualization [41]. ZINC Database serves as an openly accessible compound library containing millions of chemical structures, including NPs

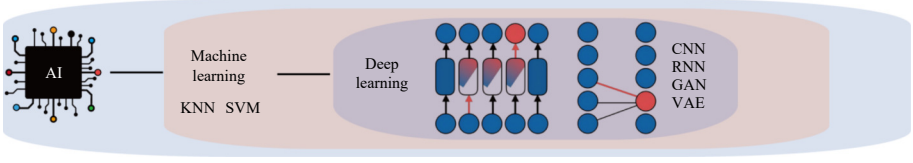

Fig. 2 Hierarchical relationship of common artificial intelligence (AI) techniques used in natural product research. Machine learning represents the core domain of AI, while deep learning is a specialized subfield within machine learning, often utilized for its powerful data pattern recognition capabilities in chemical and biological datasets.

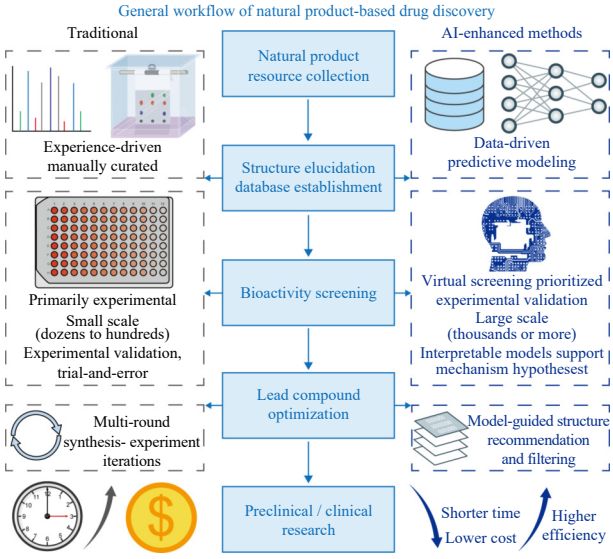

Fig. 3 General workflow and comparative integration of traditional and AI-based approaches in natural product drug discovery. Natural product drug discovery generally follows a multi-step process that includes: Natural product resource collection; Structure elucidation and library construction; Bioactivity screening; Lead compound optimization; Preclinical and clinical research. In this process, traditional and AI-enhanced approaches differ substantially. Overall, the integration of AI into natural product research makes the drug discovery process faster, cheaper, and more efficient.

and derivatives, commonly employed in virtual screening and drug discovery [42]. Minimum Information about a Biosynthetic Gene Cluster (MIBiG) functions as a standardized repository of biosynthetic gene clusters, offering detailed information on NP biosynthesis to enhance data standardization and facilitate retrieval [43].

HypoRiPPAtlas employs ML to predict and identify ribosomally synthesized and post-translationally modified peptides (RiPPs) from microbial and plant genomes, advancing NP discovery and biosynthetic pathway exploration [44]. Natural Products Magnetic Resonance Database (NP-MRD) represents an extensive open-access resource containing NMR data for over 41 000 NPs and bio-derived chemicals, supporting structural elucidation, dereplication, and characterization of NPs and metabolites [45]. SuperNatural 3.0 operates as an open-access database encompassing 449 058 natural compounds with properties, biological activity, and potential therapeutic applications, including antiviral, anti-bacterial, and anti-cancer activities, emphasizing prediction of compounds with low-calorie sweetener potential [46].

A categorized and updated summary of the general databases, as well as specialized repositories dedicated to NP research, is provided in Table 2.

2.2. Challenges faced by public repositories and strategies

Public repositories for NP research, while providing essential resources, encounter several significant challenges. The absence of effective incentive mechanisms frequently inhibits data sharing among researchers and institutions, impeding the accumulation and preservation of NP-related information. Additionally, the diverse nature of NP data, encompassing chemical structures, biological activities, metabolic pathways, and genetic information, presents considerable challenges. This diversity frequently results in format inconsistencies, nomenclature discrepancies, and compatibility issues across databases, thereby increasing data management complexity and making integration processes both time-intensive and susceptible to errors.

Current databases predominantly concentrate on individual data types, lacking comprehensive cross-disciplinary integration. Chemical data, genomic information, and metabolomic data are typically stored in separate systems, creating obstacles for researchers conducting comprehensive analyses. The enhancement of public repositories for NPs necessitates addressing key issues including data standardization, sharing incentives, and storage process optimization.

To overcome these obstacles, the promotion of open data policies and support for dataset construction adhering to the findable, accessible, interoperable, and reusable (FAIR) principles can minimize copyright and intellectual property constraints, thereby enhancing data accessibility [47]. Researchers are developing automated image recognition and ML techniques to address the challenges of converting spectra and images into data, enabling structural information extraction from images and PDF files with reduced manual intervention [48](Fig. 4).

3. Application of AI in component analysis and activity prediction of natural medicines

Natural medicines comprise hundreds or thousands of chemical components derived from diverse natural sources, including plants, animals, and microorganisms. These medicines contain highly complex active components, involving numerous compounds that potentially exhibit varied biological activities and affect multiple targets and biological pathways [49, 50]. Traditional methods for active ingredient screening primarily utilize high-throughput screening (HTS) and liquid chromatography-MS (LC-MS) techniques due to this complexity in both components and mechanisms of action [51].

Natural medicines present distinct analytical challenges compared to synthetic drugs due to their complex and diverse compositions. While synthetic drugs typically contain a single active ingredient or limited components, natural medicines often incor-

Table 1 Applications of AI in Accelerating Drug Discovery from Natural Products.

Disease	AI method	Key findings	Ref.
COVID-19	ML-based classification	A machine-learning-enhanced screening pipeline was developed to identify natural inhibitors targeting the spike receptor-binding domain (RBD) of SARS-CoV-2. The model screened 125 natural ligands and revealed several high-affinity binding phytochemicals with favorable drug-likeness profiles.	216
Malaria	Transcriptomic ML model (cross-platform integration)	A machine learning model trained on Plasmodium falciparum transcriptomic data predicted artemisinin resistance across both *in vivo* and *in vitro* platforms. The model identified conserved gene expression biomarkers and achieved top ranking in the DREAM Malaria Challenge.	217
Cancer (Microtubule target)	Deep Neural Network	Identified Eleutherobin, bruceine D, and phorbol 12-myristate 13-acetate as potent β-tubulin inhibitors	218
Parkinson's disease (PD)	Deep Learning	Identified sclareol as selective Cav1.3 blocker; validated *in vitro* and in PD mouse model	219
Type 2 diabetes (JNK1)	ML (SVM, RF, ANN)	Tricin identified from 4112 NPs; IC_{50} 17.68 μmol·L^{-1} against JNK1	220
Chagas disease	ML (RF, kNN)	Identified andrographolide (IC_{50} 2.9 μmol·L^{-1} *vs* amastigotes); selectivity comparable to benznidazole	221

Table 2 Examples of databases for natural products research.

Type	Name	Website	Last updated	Ref.
Chemical Structure and Property Databases	ChEBI	https://www.ebi.ac.uk/chebi/init.do	2025	38
	SuperNatural 3.0	http://bioinf-applied.charite.de/supernatural_3	2022	46
	PubChem	https://pubchem.ncbi.nlm.nih.gov/	2025	222
	ZINC	https://zinc15.docking.org/	2025	223
Genomic and Biosynthetic Databases	MIBiG	https://mibig.secondarymetabolites.org/	2023	43
	HypoRiPPAtlas	https://github.com/mohimanilab/seq2ripp	2022	44
Spectral and Structural Analysis Databases	NP-MRD	https://np-mrd.org	2025	45
Metabolic and Biological Function Databases	KEGG	http://www.genome.ad.jp/kegg/	2025	224
Natural Product Databases	Natural Products Atlas	https://www.npatlas.org	2024	26
	COCONUT	https://coconut.naturalproducts.net/	2025	225
	SWMD	https://swmd.co.in/	2020	226

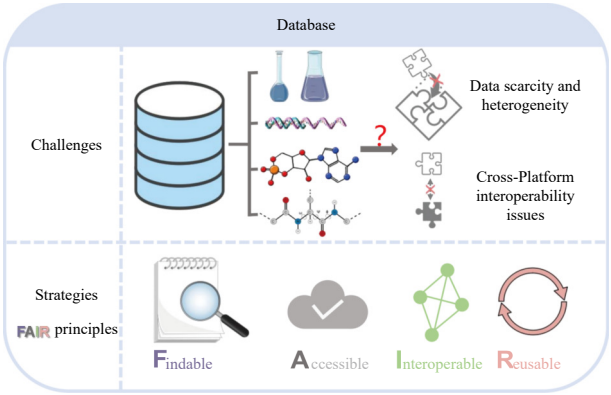

Fig. 4 Challenges and strategic solutions associated with public repositories of natural product data. Key challenges include data scarcity, heterogeneity, and cross-platform interoperability issues. To address these, implementation of the FAIR principles—Findable, Accessible, Interoperable, and Reusable—provides a systematic approach to improving data quality and reusability in the field.

porate hundreds or thousands of bioactive compounds with varying chemical structures, bioactivities, and mechanisms of action [52]. This complexity complicates the identification and isolation of therapeutically active components and their bodily interactions [52, 53]. Furthermore, NPs may exhibit synergistic or antagonistic component interactions, adding complexity to their pharmacological profile determination [54, 55].

The inherent complexity of NP molecular structures, including multi-ring skeletons, high stereochemistry, and extensive functional group modifications, poses significant challenges for traditional computational methods such as QSAR modeling and classical virtual screening, often resulting in limited throughput and accuracy [56]. AI methods offer a robust alternative by automatically learning complex structural patterns, substantially improving NP drug discovery efficiency. Traditional analytical techniques, while powerful, face considerable limitations in analyzing complex NP mixtures [57]. These methods typically require extensive sample preparation, involve high costs, and may overlook low-abundance yet pharmacologically significant compounds [58]. Moreover, they inadequately address the intricate relationships between compound chemical structures and biological activities, particularly with large, diverse datasets [59, 60, 61].

The advancement of AI technologies, particularly ML and DL, has revolutionized natural medicine component analysis and activity prediction [62, 63]. These technologies efficiently process high-dimensional, noisy, and heterogeneous data. ML algorithms, trained on extensive molecular structure and biological activity datasets, can identify hidden patterns that may elude conventional analysis [64, 65]. DL models, especially multi-layer architectures like convolutional neural networks (CNNs) and GNNs, automatically extract complex structural features without predefined descriptors [66, 67]. These capabilities enable AI methods to rapidly predict bioactivity, prioritize key components, and reduce experimental requirements, offering an effective alternative to traditional complex chemical mixture analysis methods in natural medicines.

AI efficiently processes complex drug data, extracting valuable insights from extensive activity data and chemical components, automatically learning patterns, and predicting and screening drug active components. These capabilities significantly enhance both drug development efficiency and accuracy [63, 68].

3.1. Application of ML in activity component prediction

ML represents a data-driven methodology that utilizes large volumes of existing data to train models for predicting outcomes in new, unseen data. In NP research, ML primarily serves to train models with historical activity data to predict the biological activity of unknown compounds [69].

In the component analysis of natural medicines, supervised learning algorithms in ML, such as SVM, RF, and gradient boosting machines (GBM), are predominantly employed. These algorithms incorporate biological activity data from known NP compounds to construct predictive models. For instance, Loc Nguyen and colleagues developed an integrated computational framework called iANP-EC, which combines ML techniques with evolutionary computation. The research employed four learning algorithms (k-NN, SVM, RF, and XGB) and four molecular representation schemes to construct a series of classifiers. The highest-performing four classifiers were selected and combined to form an ensemble classifier. Additionally, particle swarm optimization (PSO) optimized the weights for combining these classifiers. The model was developed using a curated dataset of 997 compounds from the NPACT and CancerHSP databases. The results demonstrated that iANP-EC is a stable, robust, and efficient framework, with an AUC-ROC value of 0.9193 and an AUC-PR value of 0.8366. A comparative analysis of molecular substructures between natural anti-carcinogens and non-anticarcinogens revealed several key substructures potentially responsible for anti-cancer activity. The research team subsequently deployed the ensemble model as an online web server with a user-friendly interface, supporting the research community in identifying NPs with anti-cancer po-

tential [70].

Another example involves the ML-based prediction of antimicrobial compounds. Many natural plant extracts demonstrate antimicrobial activity, but identifying potential anti-microbial molecules from hundreds of plant compounds presents a considerable challenge. Roberta Astolfi, utilizing experimental results and literature data from the AI4EssOil database, compiled a dataset of 82 essential oils (EOs) along with their known minimum inhibitory concentration (MIC) values. The study employed ML classification algorithms, including Support Vector Machines, Random Forest, Gradient Boosting, Decision Trees, and k-Nearest Neighbors, to establish quantitative composition-activity relationship models. Feature importance analysis, based on the Skater methodology, identified key chemical components that influence EO activity. Individual chemical components such as limonene, eucalyptol, *α*-pinene, linalool, *β*-caryophyllene, nerol, *β*-pinene, neral, and carvacrol were identified as essential for biological efficacy. The predictive performance of the ML models was validated using a test set of freshly extracted and chemically characterized EOs, achieving a prediction accuracy of 91% for new EO samples. Additionally, a strong correlation emerged between the predicted feature importance and experimental inhibition values for six selected pure compounds (limonene, eucalyptol, α-pinene, linalool, carvacrol, and thymol). Furthermore, the ML approach was extended to cytotoxicity data for 61 EOs using 3T3-Swiss fibroblasts. The analysis revealed that through blending or selectively enriching these key components, EOs with both high antibacterial activity and low cytotoxicity could be designed. These findings establish a foundation for the biological standardization of EOs and support their rational design and optimization for clinical applications [71].

This demonstrates a significant advantage of utilizing ML for activity prediction in natural medicines: it enables rapid screening of potential active components, substantially enhancing drug screening efficiency. Researchers can utilize these predictions to minimize the randomness and high costs associated with experimental processes, thereby optimizing drug development efficiency.

3.2. DL and structure-activity relationship (SAR)

DL, a fundamental branch of ML, distinguishes itself through its ability to automatically extract features from large datasets using multi-layer neural networks, enabling the recognition of more complex patterns [72]. In the component analysis and activity prediction of natural medicines, DL presents significant advantages over traditional ML methods, particularly when analyzing complex molecular structures. It automatically extracts relevant features from molecular structures, thus overcoming the limitations of traditional approaches in SAR analysis [73].

Traditional activity prediction methods typically rely on fundamental chemical features of molecules, such as molecular weight, polarity, hydrophobicity, and electron distribution. While these features provide some predictive value, they often fail to capture the subtle factors influencing activity in complex NP molecules [51]. In contrast, DL algorithms like deep neural networks (DNNs) and CNN perform end-to-end learning on molecular structures through multi-layer architectures, enabling automatic extraction of high-level features and identification of non-linear relationships between molecules and biological activity [74].

Heba Askr et al. developed an innovative model combining DL, fuzzy rough set theory, explainable AI (XAI), and medicinal chemistry to identify potential new drug candidates by analyzing experimental data from Vidarabine. The model's predictions were validated through subsequent anti-cancer experiments, leading to the identification of sulfur and magnesium oxide as drugs with significant anti-cancer activity, economic feasibility, and widespread availability. This research provides novel insights and approaches for drug repurposing research [75].

DL demonstrates significant potential when trained on pharmacological datasets, enhancing SAR predictions while determining key pharmacokinetic [absorption, distribution, metabolism, excretion, and toxicity (ADMET)] properties, including metabolic pathways, drug efficacy, and human body half-life. Notably, Dong Chen et al. developed the Algebraic Graph-Assisted Bidirectional Transformer (AGBT) framework, which combines algebraic graph and bidirectional transformer representations with various ML algorithms, including decision trees, multitask learning, and DNNs, for molecular property prediction. The AGBT framework underwent validation across eight molecular datasets, incorporating quantitative toxicity, physical chemistry, and physiology data [76]. This framework substantially enhances screening efficiency and provides essential support for decision-making in NP drug development, particularly during initial screening phases.

A notable implementation of DL appears in anti-cancer drug screening. Yu-Chiao Chiu and colleagues developed DeepDR, a DL model for predicting drug responses based on mutation and gene expression profiles of cancer cells or tumors. The model incorporates three DNNs: a mutation encoder pre-trained using a comprehensive pan-cancer dataset (The Cancer Genome Atlas, TCGA) to extract core representations from high-dimensional mutation data, a pre-trained gene expression encoder, and a drug response prediction network integrating the previous two subnetworks. Their model predicted drug responses in 9059 tumors across 33 cancer types. The model effectively identified both established drugs (such as EGFR inhibitors in non-small cell lung cancer and tamoxifen in ER + breast cancer) and novel drug targets (such as vinorelbine for TTN-mutated tumors) [77].

Yongcui Wang et al. developed DeepDRK for drug response analysis. Their research involved training DNNs using data from over 20 000 pan-cancer cell line-anticancer drug pairs, facilitating information transfer between different drugs and cancer types. The data utilized similarity matrices generated through kernel methods, incorporating multi-omics data from various sources, including genomics, transcriptomics, epigenomics, chemical properties, and known drug-target interactions. The model demonstrated high accuracy and stability when validated on newly established patient-derived cancer cell lines. Significantly, DeepDRK's predictions showed strong correlation with clinical patient outcomes and identified several potential drug repurposing candidates [78]. These DNN-based tools, trained on extensive datasets, have identified several novel natural compounds with anti-cancer properties previously undetected through traditional HTS.

4. Application of AI in the pharmacological mechanism and multi-target analysis of natural medicines

Natural medicines exhibit complex pharmacological mechanisms involving multiple targets and biological pathways [79]. Unlike conventional single-target drugs, natural medicines typically modulate biological processes through multiple targets and pathways, producing more complex and diverse therapeutic effects [16]. For instance, numerous plant-derived NPs demonstrate anti-inflammatory and anti-oxidant effects while potentially influencing the immune system, nervous system, or metabolic pathways [80-82]. Several applications of AI in the pharmacological mechanisms and multi-target analyses of natural medicines are listed in Table 3. Understanding these comprehensive pharmacological mechanisms remains essential for advancing clinical applications and facilitating new drug development.

The evolution of AI technologies, particularly the integration of ML and DL methods, has revealed significant potential in phar-

macological mechanism and multi-target analysis [83, 84]. Through the integration of various "omics" data, including genomics, proteomics, and metabolomics, AI enables the discovery of natural medicine targets and mechanisms of action in a high-dimensional context, providing a more comprehensive and accurate representation of drug actions [85, 86].

4.1. Application of AI in the pharmacological mechanism analysis of natural medicines

Natural medicines typically involve multiple molecular-level interactions in their pharmacological actions, and conventional experimental methods face challenges in revealing these complex mechanisms efficiently. AI, particularly ML, analyzes large-scale "omics" data—including gene expression profiles, protein-protein interaction (PPI) networks, and metabolite data—to derive valuable insights and identify the targets and pharmacological effects of natural medicines [87]. AI can identify potential effects of natural medicines on specific diseases by analyzing gene expression changes in treated cells or tissues. These genes may function as direct targets or represent key pathways regulated by drug intervention. Beyond gene expression analysis, AI integrates PPI networks to identify critical signaling pathways influenced by natural medicines. Through large-scale PPI network analysis, AI identifies potential target proteins and predicts their biological process modulation. Additionally, AI serves a crucial role in metabolomics analysis, examining natural medicine metabolic products and elucidating their biotransformation pathways and metabolic dynamics [88-90].

A significant challenge in NP research lies in the often unclear or inadequately characterized efficacy of numerous natural medicines. Traditional methods frequently prove insufficient in fully capturing or verifying their therapeutic effects due to the complexity of their multi-component and multi-target actions, resulting in uncertainties regarding their clinical application and acceptance [91, 92]. AI presents robust solutions for systematically elucidating the efficacy of natural compounds. Through the integration of diverse data sources—including clinical trial outcomes, real-world patient records, multi-omics data, and literature mining—AI models can detect therapeutic benefit patterns, reveal hidden connections between compounds and disease pathways, and predict natural medicines' clinical efficacy with enhanced precision [93, 94]. Additionally, XAI approaches facilitate the visualization and interpretation of compound pharmacological actions, thereby connecting molecular mechanisms with observable therapeutic outcomes [95, 96]. These capabilities enhance understanding of natural medicines' efficacy while accelerating their rational development and evidence-based clinical translation [97].

Miaobo Ye et al. investigated the mechanism of Toujie Quwen Granules in treating coronavirus disease 2019 (COVID-19) pneumonia through network pharmacology, molecular docking, and integrated surface plasmon resonance (SPR) technology. The researchers integrated chemical constituents and target information from various Chinese medicinal herbs to establish a Chinese medicine-compound-target network. Using Metascape, they conducted PPI network analysis, Gene Ontology (GO) functional enrichment, and KEGG pathway enrichment analysis on core targets to elucidate their mechanisms of action. Molecular docking analysis examined interactions between the top 34 compounds and key SARS-CoV-2 enzymes, including 3CL protease and ribonucleic acid (RNA)-dependent RNA polymerase (RdRp). The researchers then evaluated 13 compounds with the lowest affinity scores for interactions with ACE2, SARS-CoV-2 spike protein, and interleukin-6 (IL-6). SPR experiments validated the binding affinity and biological activity of selected compounds, including quercetin, astragaloside IV, rutin, and isoquercitrin. These results provide substantial evidence and theoretical support for Chinese medicine's potential mechanisms against SARS-CoV-2 [98].

Qin Deng et al. investigated the mechanism of *Andrographis paniculata* against solar dermatitis utilizing network pharmacology, molecular docking, and experimental validation. The study identified active components and potential targets of *Andrographis paniculata* from TCMSP and Swiss Target Prediction databases. Potential therapeutic targets for solar dermatitis were gathered from GeneCards, DrugBank, and OMIM databases. The researchers constructed PPI networks and compound-target-disease (C-T-D) networks using Cytoscape to identify key therapeutic targets in solar dermatitis, including AKT-1, TNF-α, IL-6, MMP9, EGFR, and PTGS2. GO functional analysis and KEGG pathway enrichment analysis using the DAVID database indicated the PI3K-Akt signaling pathway's central role in *Andrographis paniculata*'s therapeutic effects on solar dermatitis. Molecular docking assessed binding affinity between key components and target genes. Using UVB-irradiated HaCaT keratinocytes and UVB-irradiated ICR mice as models, they confirmed that *Andrographis paniculata* demonstrates significant anti-inflammatory and skin repair effects through the PI3K-Akt signaling pathway [99].

ML and DL have gained recognition for their contributions to pharmacological mechanism analysis [100]. Although these AI technologies are frequently used interchangeably, they demonstrate distinct characteristics in natural medicine research applications. ML techniques, including SVM and RF, typically utilize pre-processed, structured data requiring manual feature selection. These methods prove effective when relationships between variables are relatively straightforward and datasets are well-defined [97, 101]. Conversely, DL methods, such as CNNs and recurrent neural networks (RNNs), demonstrate excellence in analyzing unstructured, high-dimensional data, including genomics, proteomics, and metabolomics [102]. DL models automatically extract complex features from data without manual intervention, making them particularly valuable for large-scale and heterogeneous datasets.

Network-based AI methods, particularly GNNs, allow researchers to model complex relationships between molecules, targets, and biological pathways [103]. This approach provides a comprehensive view of pharmacological mechanisms while improving predictive accuracy for multi-target interactions. Unlike traditional AI techniques focusing on single-target drug effects, modern ML/DL algorithms identify subtle, non-linear relationships between multiple natural medicine components, substantially enhancing understanding of their therapeutic potential [104, 105]. These developments establish AI as an essential tool in pharmacological research, enabling more precise and efficient mapping of complex molecular networks involved in natural medicine's actions [106].

4.2. Application of AI in multi-target drug design

The intricate composition and multi-target mechanisms of

Table 3 Applications of AI in the Pharmacological Mechanism and Multi-target Analysis of Natural Medicines

Tool	Description	Availability/Website	Ref.
TIGER	Reveal the target promiscuity of pharmacologically active compounds	Not disclosed	192
STarFish	A stacked Ensemble Target Fishing Approach	https://github.com/ntcockroft/STarFish	227
SPiDER	Identifying the macromolecular targets of de novo-designed chemical entities through self-organizing map consensus	http://modlab-cadd.ethz.ch/software/	228

natural medicines have established them as a crucial focus in contemporary drug development. In contrast to conventional single-target drugs, natural medicines modulate biological systems through multiple targets and pathways, presenting significant therapeutic potential across diverse diseases [107]. These medicines comprise multiple bioactive components, each capable of modulating various biological pathways and cellular processes. For example, numerous plant-derived compounds demonstrate multiple therapeutic effects in cancer treatment, targeting various mechanisms including inflammation, cell cycle regulation, oxidative stress response, and apoptosis, demonstrating particular effectiveness against cancer, cardiovascular diseases, metabolic disorders, and other conditions [108]. Conversely, traditional single-target drugs typically address one specific target, potentially limiting their therapeutic efficacy [109]. Consequently, the development of multi-target drug design has become increasingly critical in natural medicine research and development.

AI technologies, particularly ML and DL, serve a crucial role in investigating the multi-target actions of natural medicines through the analysis of large-scale multi-omics data (including genomics, proteomics, and metabolomics). AI facilitates the identification of potential targets for various natural medicine components, elucidating their effects on multiple biological pathways and enabling the development of multi-therapeutic drugs. The three primary methods of AI applications in this process comprise: ligand-based methods, structure-based methods, and ligand-and-target networks. Ligand-based methods utilize drug molecule (ligand) characteristics for target prediction, fundamentally based on ligand similarity and chemical fingerprints (Fig. 5A). Structure-based methods employ three-dimensional structural information of the target for prediction, typically utilizing computational simulations, such as molecular docking and molecular dynamics simulations, to predict binding modes and affinities between drug molecules and targets (Fig. 5B). Ligand-target networks establish network models between ligands and targets to reveal complex relationships among ligands, targets, and biological pathways, integrating both ligand-based and structure-based approaches for network-based analysis [110,111](Fig. 5C).

In drug-target prediction, AI initially processes large-scale drug-target databases to develop models correlating drugs with their targets [112]. Through training on natural medicine compound data, AI identifies potential target interactions [113]. Furthermore, AI has enhanced the integration of multi-target drug design with network pharmacology, examining multi-pathway and multi-level drug effects through layered drug-target-disease networks [114]. AI analyzes gene, protein, and interaction networks to identify multiple targets of natural medicines and predict their influence on various biological pathways [115]. Through virtual screening, AI identifies candidate molecules exhibiting high affinity, favorable pharmacokinetic properties, and low toxicity, substantially improving drug development success rates [116,117]. This methodology enhances drug design precision while reducing experimental costs and duration. AI-driven target predictions enable researchers to select natural medicine components more effectively, facilitating advances in multi-target drug design.

Through the application of AI technologies, researchers can develop more effective multi-target drugs that simultaneously target multiple disease-related pathways and regulate various biological processes, resulting in improved therapeutic outcomes. For example, in a study by Huaguo Liang and colleagues, the therapeutic potential of Selaginella doederleinii Hieron (S. doederleinii) was investigated for treating nasopharyngeal carcinoma (NPC) —a condition characterized by complex etiology and late-stage diagnosis. The researchers identified NPC-related target genes through databases such as GeneCards, OMIM, and DisGeNET, and conducted weighted gene co-expression network analysis (WGCNA) on the GSE53819 dataset, revealing several

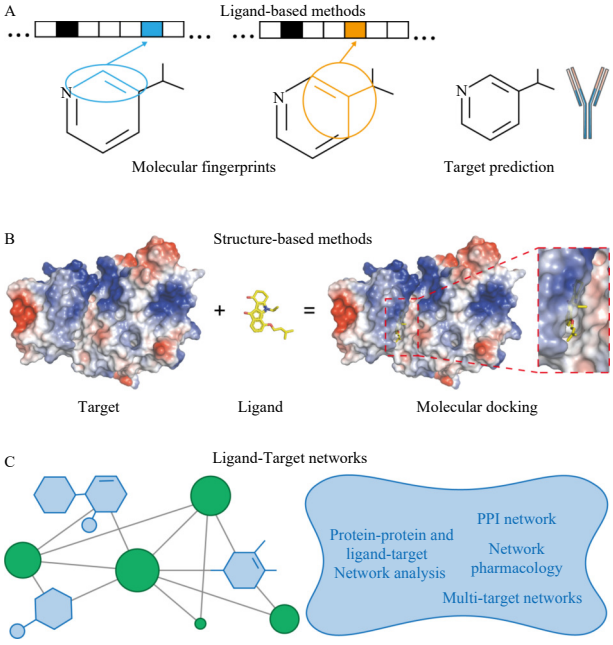

Fig. 5 Common AI-based approaches of predicting targets for natural products. (A) Ligand-based methods utilize the intrinsic chemical features of small molecules (ligands) to predict potential targets through similarity comparisons and molecular fingerprints. (B) Structure-based methods rely on the three-dimensional structural information of biological targets and use techniques such as molecular docking and molecular dynamics simulations to estimate binding modes and affinities. (C) Ligand-target network models integrate both ligand-based and structure-based data to construct multi-target interaction networks, thereby elucidating complex relationships between compounds, targets, and biological pathways.

significant gene modules associated with NPC. Active components and their targets from *S. doederleinii* were identified from the TCMSP and other databases, revealing 32 overlapping genes. GO and KEGG pathway analyses identified key biological processes, including protein phosphorylation and cell cycle regulation. A PPI network was constructed, with the cytoHubba tool identifying six critical genes (BCL2, MAPK14, ABCB1, PLK1, ATM, and HMOX1) that were closely related to NPC prognosis and immune microenvironment, based on Kaplan-Meier survival analysis and immune infiltration studies. Additionally, single-cell RNA sequencing was employed to examine the expression patterns of these key genes across different immune cell types, and their roles in malignant cell differentiation were explored using pseudotime trajectory analysis. Molecular docking and dynamics simulations further confirmed the strong binding affinity and stability of the Berberine-MAPK14 and Matairesinol-PLK1 complexes, suggesting their potential therapeutic value. Overall, the study indicates that the active components of S. doederleinii may substantially contribute to NPC treatment through synergistic multi-pathway and multi-target effects [118].

In conclusion, AI technology serves a vital role in multi-target drug design for natural medicines. Through AI implementation, researchers can elucidate the multi-target action mechanisms of natural medicines and expedite drug screening and optimization processes, facilitating the development of new natural drugs. As AI technology advances, the field of multi-target drug design in natural medicines will enter a new phase of precision and efficiency.

5. Application of AI in predicting toxicity and side effects of NPs

Although NPs are generally regarded as having high safety profiles due to their origins in nature, they can still present toxicity or side effects, particularly when used in large doses or over

extended periods. For example, Gardenia jasminoides J. Ellis (Gardeniae Fructus, GF), a commonly used herbal medicine, is known for its hepatoprotective effects, making it an important component in the treatment of various liver diseases. However, long-term use or overdose of GF may lead to hepatotoxicity, raising concerns regarding its clinical application [119, 120]. Therefore, despite the broad clinical applications of NPs, their safety must still be carefully evaluated. Traditional toxicity prediction methods often rely on animal testing, which involves extensive experiments on animals to assess the toxicity of drugs or their chemical components. However, animal testing is not only time-consuming and costly (Fig. 6A) but also raises significant ethical concerns [121, 122] (Fig. 6B). Moreover, animal models do not always accurately reflect human toxicity responses, posing challenges for reliable toxicity prediction [123, 124] (Fig. 6C). In this context, AI technologies, particularly ML and DL, have emerged as essential tools for predicting the toxicity and side effects of NPs [125]. By analyzing large toxicity datasets, AI can assist researchers in predicting the potential toxicity of NPs through data modeling and pattern recognition. This approach helps reduce the failure rates in preclinical and clinical trials and provides a theoretical basis for optimizing drug development [126].

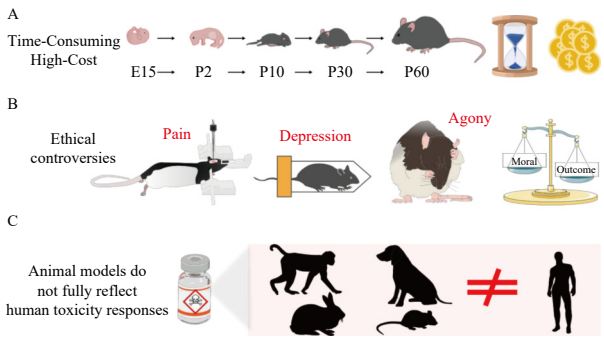

Fig. 6 Limitations of traditional toxicity assessment methods. (A) Animal testing is time-consuming, labor-intensive, and expensive. (B) Ethical concerns raise, as laboratory animals are often subjected to suffering, distress, and aggressive behavioral changes. (C) Animal models do not always accurately reflect human physiological and toxicological responses, limiting the translatability of the results.

The implementation of AI technology effectively addresses the limitations of traditional toxicity prediction methods. Through analysis of extensive toxicity databases, AI can identify correlations between molecular structural features and toxicity profiles. These features, encompassing molecular weight, chemical structure complexity, functional group characteristics, and additional parameters, are fundamental determinants of toxicity. Through training DL models, AI can automatically identify these features and utilize them to predict the potential toxicity of novel compounds [127]. Furthermore, AI's contribution to toxicity prediction extends beyond structural analysis. Research has demonstrated that drug toxicity is intrinsically linked to metabolism, particularly involving drug-metabolizing enzymes such as the CYP450 family. These enzymes can generate toxic metabolites within the body, potentially leading to adverse reactions. Consequently, AI can integrate drug metabolism characteristics with ADMET properties, enhancing prediction accuracy [128-130]. By simulating drug metabolism pathways, AI can forecast metabolic outcomes and identify potentially toxic metabolites. Beyond considering molecular structures and metabolic characteristics, AI can synthesize toxicity data from multiple sources for comprehensive analysis [131, 132]. This multifaceted approach substantially improves the reliability and precision of toxicity predictions, advancing safer drug development practices.

AI technology has become extensively employed in predicting the toxicity and side effects of NPs. For instance, Qi Yang et al. developed a ML-based model for predicting hepatotoxicity and validated it using 56 chemical constituents of Gardenia jasminoides. The model's effectiveness was confirmed through literature reviews, principal component analysis (PCA), and structural comparison techniques. Their findings revealed that the hepatotoxic constituents of Gardenia exhibit dual effects—at certain doses, they can display both therapeutic and toxic properties [133]. Another significant example comes from Chen Jia et al., who developed an effective ML model to predict the toxicity of traditional Chinese medicines (TCMs). They correlated the toxicity of TCMs with analytical descriptors derived from electron ionization MS (EI-MS) data. The model achieved a classification accuracy of over 0.74, enabling the identification of specific toxic components. Furthermore, molecular dynamics simulations were employed to investigate how these components interact with key protein targets, such as hepatic cytochrome P450 3A4. This approach provides valuable insights into the toxicological profiles of TCMs, facilitating the maximization of therapeutic benefits while minimizing adverse effects. Additionally, it supports the prediction of toxicity in unknown mixtures found in real-world environments [126] (Fig. 7A).

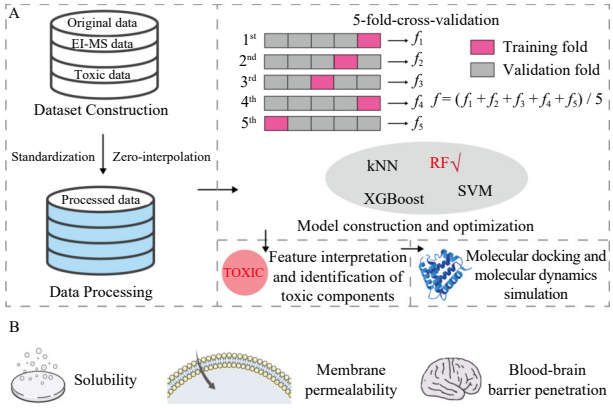

Fig. 7 Application of AI in predicting toxicity and improving bioavailability. (A) Workflow for predicting the toxicity of traditional Chinese medicines (TCMs) using EI-MS and machine learning.The process begins with data collection and preprocessing: TCM samples are selected from the Chinese Pharmacopoeia and subjected to electron ionization mass spectrometry (EI-MS). The spectral data undergo zero-filling interpolation, feature cleaning, and intensity normalization. Next, four machine learning algorithms—Random Forest (RF), XGBoost, Support Vector Machine (SVM), and k-Nearest Neighbors (kNN)—are trained, and their performance is evaluated using five-fold cross-validation. The RF model is selected as the optimal performer. Feature importance analysis from the RF model is then used to identify potentially toxic compounds. Finally, molecular dynamics simulations are employed to investigate the interaction mechanisms between the identified toxic compounds and relevant proteins. (B) Artificial intelligence techniques are used to predict and optimize key pharmacokinetic and physicochemical properties, including solubility, membrane permeability, and blood-brain barrier penetration to improve drug bioavailability and safety.

In addition to toxicity prediction, AI technologies serve a crucial role in optimizing the bioavailability and safety profiles of NPs [134, 135]. Bioavailability, which refers to the extent and rate at which an active ingredient is absorbed and becomes available at the site of action, is a critical factor affecting a compound's clinical efficacy and safety. Poor bioavailability can result in increased dosing requirements and a higher risk of adverse effects [136]. AI models, particularly those based on DL and reinforcement learning, have been employed to predict and optimize physicochemical properties such as solubility, membrane permeability and blood-brain barrier penetration, all of which are closely associated with bioavailability and safety [137-139](Fig. 7B). For example, Liu et al. developed and validated advanced AI models—including RF, Gradient Boosting Regression Trees (GBRT), and Extremely Randomized Trees (ERT)—to predict the solubility of the anti-cancer drug busulfan in a supercritical CO_2 system. Among the tested models, GBRT achieved the best performance, with an RMSE of 1.03×10^{-4} and MAPE of 0.362, demonstrating

high predictive accuracy ($R^2 > 0.9$). Furthermore, the model identified an optimal solubility condition at 38.3 °C and 333.1 bar, yielding a solubility of 1.36×10^{-3} mol fraction. This study highlights the potential of AI to support early-stage formulation design and solubility optimization, which are critical for improving the bioavailability and delivery of natural and synthetic compounds [140].

Overall, the application of AI technology in predicting the toxicity and side effects of NPs has substantially enhanced the efficiency and accuracy of toxicity predictions, significantly reducing risks in the drug development process. Through the integration of extensive toxicity data, analysis of chemical structure and metabolic characteristics of drug molecules, and combination of toxicity databases for pattern recognition, AI assists researchers in identifying potential safe drug candidates. This reduces dependence on traditional experimental methods, decreasing both experimental costs and time. Additionally, AI's capabilities in DL and molecular simulations enable more precise safety assessments, providing a solid scientific foundation for the clinical application and market promotion of NPs. With ongoing advancements in AI, the future of toxicity prediction for NPs is positioned to become increasingly efficient and accurate, heralding a new era for NP development.

6. Challenges and strategies for further integrating AI with NP research

As AI continues to be integrated into NP research, significant advancements have emerged, including the optimization of NP ingredient screening, toxicity prediction, and pharmacological mechanism analysis. Despite these promising achievements, the application of AI in this field still faces several challenges. Addressing these obstacles will be crucial for the broader adoption and success of AI technology in NP research moving forward. A summary of the challenges and strategies for further integrating AI into NP research is presented in Table 4.

Table 4 Challenges and strategies for further integrating artificial intelligence with natural products research.

Challenge	Strategies
Data Quality and Diversity	Improve and expand databases
	Enhance collaboration with regulatory agencies and big data platforms
	Promote data integration and sharing
Model Explainability	Develop Explainable AI (XAI) methods, such as attention mechanisms and feature visualization, to enhance transparency.
Interdisciplinary Collaboration	Foster cross-disciplinary platforms and partnerships among academia, pharmaceutical companies, and AI startups.
Hybrid Models with Experimental Validation	Build closed-loop systems combining AI prediction with experimental feedback to iteratively refine drug discovery
Ethical Concerns	Implement data protection measures
	Establish AI-specific ethical frameworks
	Encourage interdisciplinary ethical review
Cost and Infrastructure	Provide targeted funding support
	Establish shared computational platforms
	Encourage cross-institutional collaborations to share costs and resources

6.1. Data quality and diversity

The effectiveness of AI models is fundamentally dependent on high-quality, diverse, and standardized data. In NP research, several significant obstacles impede the successful implementation of AI technologies, including inadequate data standardization, limited partnerships with drug regulatory agencies and big data companies, and insufficient data volume [141, 142]. NPs exhibit complex and diverse compositions, frequently featuring large, non-standardized chemical structures. While databases such as ChEMBL, PubChem, and Tox21 have accumulated substantial data in specific domains [143, 144], biological activity data for NPs originate from diverse experimental methodologies, potentially compromising comparability. This fragmentation creates challenges including data scarcity, imprecise labeling, and irregular updates, hindering effective cross-source training for AI models [145, 146].

Moreover, numerous NP components remain unidentified or lack standardization, creating obstacles for researchers attempting to characterize their multiple biological activities accurately. Future research must prioritize the enhancement of natural NP databases and the establishment of more comprehensive, high-quality data systems. Additionally, emphasis should be placed on improving the integration and sharing of data from multiple sources [147]. For instance, partnerships with drug regulatory agencies and AI-driven big data platforms could utilize HTS technologies and artificially synthesized chemical substances to systematically expand NP databases. Such expansion would enhance AI models' capacity to adapt to NPs' inherent diversity and complexity [148, 149].

Training data that is biased or incomplete can significantly affect AI-driven drug discovery research in several ways. First, training data lacking representativeness may cause model overfitting within a limited chemical space, restricting the generation of compounds with novel functions and properties. Second, data bias may result in models incorrectly estimating the biological activity of certain compound classes, leading to elevated rates of false positives or false negatives. Third, given that NPs typically operate through complex multi-objective mechanisms, training data that is incomplete or narrow in scope may compromise the model's ability to accurately identify structure-target relationships.

Several advanced AI techniques can address the challenges posed by biased or incomplete data. Transfer learning facilitates knowledge reuse from related, well-annotated datasets to enhance performance in target domains with limited data, effectively addressing issues of data scarcity or skewed distributions [150]. Federated learning implements a decentralized learning approach enabling collaborative model training among institutions without raw data sharing, preserving privacy while enhancing data diversity and robustness [151]. Furthermore, synthetic data generation, utilizing generative adversarial networks

(GANs) or SMILES-based molecular generators, can supplement underrepresented data classes or simulate rare compound types, improving AI models' generalization capabilities [152]. These strategies collectively enhance the robustness, generalizability, and reliability of AI systems in NP research.

6.2. Model explainability

Despite the remarkable predictive accuracy of DL and other AI technologies in NP research, their "black box" nature presents a significant challenge [153]. While DL models effectively learn features from large datasets for predictions, their underlying decision-making processes often resist interpretation [154]. This opacity is particularly problematic in drug development, where drug safety and efficacy depend not only on prediction accuracy but also on understanding the underlying mechanisms. Clear comprehension of the prediction process is crucial for supporting clinical applications [155, 156]. Regarding AI predictions of NPs' multi-target effects, DL models may identify potential targets and pathways but frequently fail to elucidate specific interaction mechanisms, biological pathways, and potential side effects. This lack of transparency constrains AI's utility in drug development, particularly during critical decision-making phases, as it provides insufficient confidence for drug developers [97] (Fig. 8A). Addressing this challenge requires focused research on advancing XAI. The development of more transparent algorithms —including attention mechanisms and feature visualization techniques —will enable researchers to better understand AI predictions and optimize model outputs [157, 158].

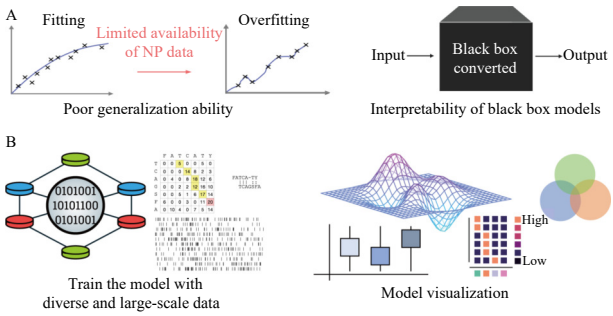

Fig. 8 Challenges and potential strategies for applying AI models in natural product research. (A) A major limitation is the lack of sufficient high-quality, annotated data on natural products, which undermines the generalization ability of AI models. Additionally, many AI models function as "black boxes", offering limited interpretability. (B) Strategies such as training models with large-scale, diverse datasets and employing model visualization techniques can improve both performance and transparency.

Ongoing initiatives aim to enhance the interpretability of AI models in NP research. These efforts seek to transform the traditional "black box" paradigm into a "glass box" through multidimensional technological strategies and comprehensive integration with chemical and biological knowledge. A fundamental approach involves designing models with built-in attention mechanisms, such as Transformers and GNN, which highlight molecular regions that the model examines during prediction, including pharmacophores or functional groups [159]. Feature attribution methods, notably SHapley additive explanations (SHAP) and local interpretable model-agnostic explanations (LIME), are extensively employed to assess individual molecular descriptors' contributions to biological activity prediction, enhancing model transparency and reliability [160-162]. Visualization techniques (such as heatmap annotations in DeepChem) and knowledge embedding methods (e.g., retaining generation constraints for active scaffolds) further align model outputs with domain-specific medicinal chemistry principles. Visualization strategies and knowledge embedding approaches continue to harmonize model outputs with specific domain chemical understanding [163]. Furthermore, the extensive adoption of open-source tools such as RDKit and ChEMBL, alongside standardized evaluation frameworks like chemical efficacy scores, promotes greater reproducibility and transparency in AI interpretability research [164] (Fig. 8B).

6.3. Emerging AI technologies in NP research

Recent years have witnessed substantial advancement in NP research through various sophisticated AI methods, particularly Transformers, diffusion models, and large language models (LLMs). Transformers effectively capture long-range dependencies through self-attention mechanisms and have facilitated the generation of extensive NP-like compound libraries, such as NPGPT, expanding chemical space beyond 67 million structures [165]. In virtual screening and target prediction tasks, Transformer-based architectures like MolTrans have markedly enhanced the prediction accuracy of NP-protein interactions [166]. Fragment-based conditional models such as NIMO enable directed optimization of biological activity while maintaining the structural diversity inherent to NP scaffolds [167]. Additionally, chemical language models have demonstrated efficacy in designing novel bifunctional ligands capable of targeting multiple proteins simultaneously [168].

Diffusion model employs forward backward noise processes to facilitate the generation and optimization of complex molecules in three-dimensional space. MD3MD, a multiscale equivariant diffusion model, captures ring systems and chiral features across hierarchical levels within NPs [169], while latent diffusion models have successfully designed diverse and effective anti-microbial peptides [170]. LLMs, trained on extensive chemical and textual datasets, demonstrate significant potential in knowledge integration and mechanistic reasoning. Models such as ChemBERTa provide comprehensive molecular embeddings that enhance property prediction and bioactivity modeling for NPs [171].

6.4. Interdisciplinary collaboration

The successful implementation of AI technology in natural medicine research relies not only on technological advancement but also on the comprehensive integration of life sciences and computer science [56]. This field encompasses multiple disciplines, including chemistry, pharmacology, biology, and clinical medicine, necessitating researchers to develop interdisciplinary expertise [172]. Pharmacological research requires understanding the correlation between compound structures and their biological activity, while computer science expertise is essential for managing large-scale chemical data and developing effective ML models [173]. Maximizing AI's potential in natural medicine research requires fostering collaboration among life sciences, computer science, and data science (Fig. 9).

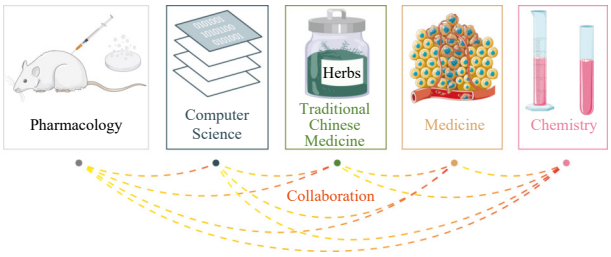

Fig. 9 Promoting interdisciplinary collaboration to advance AI-integrated natural product research. To further integrate AI into natural products research, fostering interdisciplinary collaboration across fields such as medicine, traditional Chinese medicine, computer science, pharmacology, and chemistry is a key strategy to drive innovation and address discipline-specific challenges.

Successful applications of AI in natural medicine frequently emerge from interdisciplinary collaboration [174]. This collaborative approach becomes increasingly vital as research complexity

grows in future drug development. Cross-disciplinary communication and knowledge integration serve as fundamental drivers of progress [175, 176]. Advancing interdisciplinary cooperation requires both academic and industrial sectors to establish comprehensive collaboration platforms and mechanisms [177]. Examples include developing partnerships between pharmaceutical companies and AI startups, creating shared resource platforms, and facilitating multi-expert collaboration, which accelerates AI implementation in natural medicine research and creates innovative drug discovery opportunities.

6.5. Hybrid approaches integrating AI and traditional pharmacology

While AI provides robust predictive capabilities, experimental validation remains crucial for ensuring the clinical relevance and safety of NPs. Research on hybrid approaches combining AI predictions with traditional pharmacological testing remains in its early stages. Most investigations have focused separately on computational prediction or experimental confirmation, with limited systematic integration [100, 115, 178]. Future research should prioritize developing closed-loop systems that combine computational insights with experimental feedback to refine drug discovery strategies iteratively. This methodology would utilize the complementary advantages of AI's high-throughput predictive capabilities and traditional pharmacology's empirical validation, fostering a more balanced, efficient, and reliable natural medicine development pipeline.

6.6. Ethical concerns and mitigation strategies in AI-based NP research

The increasing integration of AI in NP research has raised several ethical concerns requiring careful attention. Data privacy and security concerns emerge when AI models utilize extensive datasets containing sensitive biological or clinical information. Breaches through unauthorized access or misuse of such data can have serious ethical implications [179]. Several mitigation strategies should be implemented to address these challenges. Strict adherence to data protection regulations and implementation of secure data-sharing frameworks are fundamental [179, 180]. Additionally, developing specific ethical guidelines for AI applications in drug discovery, including natural medicine research, will foster responsible innovation. The establishment of interdisciplinary ethics review committees and promotion of open science practices can ensure that AI research in this field maintains robustness, fairness, and transparency [181, 182]. As AI transforms NP research, incorporating ethical considerations throughout model development and application remains essential for maintaining public trust and ensuring responsible AI-driven discoveries.

6.7. Cost and infrastructure: overcoming financial and logistical barriers in AI integration

Although AI technology demonstrates substantial potential in NP research, its broad implementation faces considerable financial and infrastructure-related obstacles. The implementation of AI solutions requires substantial computing resources, including high-performance hardware and specialized software tools [183]. The acquisition, maintenance, and upgrading of these resources demand significant investment, presenting a substantial barrier, particularly for academic institutions or small research organizations with limited funding. Furthermore, AI models require continuous data updates and optimization. As research advances and datasets expand, AI models must undergo regular updates to maintain predictive accuracy and adaptability. This ongoing process demands substantial computing power and skilled personnel, thereby increasing the financial requirements of AI technology implementation [184, 185].

Several approaches can address these challenges. Governments and research funding agencies should provide targeted financial support to promote AI adoption in NP research, particularly in resource-limited academic communities. Research institutions can establish collaborative shared computing platforms and open-source tools, reducing individual institutions' dependence on costly hardware and software. Furthermore, cross-institutional partnerships focused on data sharing and collaboration can help distribute AI implementation costs and minimize individual institutional financial burdens [186]. These measures can significantly reduce AI-related expenses in NP research, enabling broader participation in AI-driven research and promoting its widespread application in this field.

7. Conclusions and future perspectives

AI has emerged as a catalyst in accelerating NP drug discovery, offering capabilities beyond traditional experimental and computational workflows - particularly in managing structural complexity and analyzing large heterogeneous datasets. NPs encompass extensive chemical diversity. Traditional screening and biochemical methods face challenges including lengthy cycles, high resource requirements, and difficulties in precisely identifying active ingredients from NPs [11, 187]. AI technologies, including DL, ML, and natural language processing, efficiently extract potential structural patterns from chemical structures, multi-omics data [188, 189], and literature [190, 191], effectively prioritizing bioactive molecules, predicting targets [192], and elucidating multi-component and target mechanisms of action [193]. Additionally, AI enhances pharmacokinetics through optimization of bioavailability, stability, and selectivity, leading to improved clinical outcomes [194].

Despite AI's significant potential in NP research, challenges persist, particularly regarding natural compound complexity and the requirement for extensive experimental data. The multi-target nature of NPs and their complex mechanisms necessitate further investigation. Additionally, concerns regarding data quality, privacy protection, and data integration require attention [195, 196]. Furthermore, challenges in data quality, model interpretability, and ethical considerations have become increasingly significant. Various solutions are emerging to address these challenges. The community promotes data standardization and sharing, developing high-quality, open knowledge graphs for AI to address sample scarcity and enhance data consistency [197]. XAI is being integrated at the algorithm level to ensure traceable and verifiable model outputs [198]. Human-machine collaborative verification, ethical guidelines, and regulatory frameworks clarify intellectual property and biological resource utilization compliance for AI-generated molecules, ensuring fair and sustainable research outcomes [199]. Data augmentation, synthetic spectrum generation, transfer learning, and other methodologies expand sample sizes while controlling bias [200]. These multifaceted approaches collectively promote reliable, transparent, and responsible AI application in NP research.

Considering regional diversity in natural medicines, AI research should focus on areas offering immediate and significant benefits [201]. In the near term, AI should concentrate on expediting bioactive component identification from well-documented medicinal plants, particularly those with extensive ethnopharmacological records [202]. Moreover, AI should emphasize toxicity prediction and multi-target mechanism analysis, as these areas can rapidly enhance NPs' safety and therapeutic precision. This focused approach to AI-driven research can generate more direct translational outcomes, establishing a robust foundation for NPs' broader application in modern drug discovery [203].

In addition to AI, other technologies such as HTS and combinatorial chemistry have substantially enhanced natural medicine research. HTS facilitates the rapid screening of thousands of compounds, delivering high efficiency in early drug discovery [204, 205]. Combinatorial chemistry enables the generation of extensive libraries of chemical compounds, offering structural diversity for screening [206, 207]. Compared to these conventional technologies, AI delivers additional advantages through analyzing complex biological data, predicting multi-target interactions, and guiding rational drug design [208]. The integration of AI with complementary technologies such as HTS and combinatorial chemistry enhances the efficiency of NP-based drug discovery. In HTS, AI models trained on initial experimental results predict potentially active compounds, thereby filtering large compound libraries prior to physical screening, reducing cost and time [209-211]. In combinatorial chemistry applications, generative models—including deep reinforcement learning and graph-based neural networks—design virtual libraries with desired pharmacophores or scaffold constraints [212-214]. This AI-guided generation enhances the quality of synthesized compound pools by prioritizing molecular structures with high predicted bioactivity or favorable ADMET profiles. Such integration presents an effective strategy for focusing experimental resources on the most promising candidates. Future research should investigate collaborative models that combine AI-driven predictive capabilities with the experimental throughput of HTS and the diversity generation capacity of combinatorial chemistry, aiming to optimize the discovery efficiency and innovation potential of natural medicines [215].

Funding

The work was supports from the National Key Research and Development Program of China (No. 2020YFE0202200); the National Natural Science Foundation of China (Nos. 81903538, 82322073, 92253303); the Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (No. ZYYCXTD-D-202004); the Science and Technology Commission of Shanghai Municipality (Nos. 22ZR1474200, 24JS2830200). Some of the illustrations/icons used in this review were obtained from SciDraw and Bioicons, which provide open-source scientific illustrations and icons. We hereby express our sincere gratitude.

Availability of supporting information

Supporting information for this work can be obtained by contacting the corresponding authors *via* E-mail.

Declaration of competing interest

These authors have no conflict of interest to declare.

References

1 Mullowney MW, Duncan KR, Elsayed SS, et al. Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov*. 2023;22(11):895-916. https://doi.org/10.1038/s41573-023-00774-7
2 Chopra B, Dhingra AK. Natural products: a lead for drug discovery and development. *Phytother Res*. 2021;35(9):4660-4702. https://doi.org/10.1002/ptr.7099.
3 Ren Y, Zhang X, Li Y, et al. Research progress and applications of iron-based nanozymes in colorimetric sensing of agricultural pollutants. *Biosens Bioelectron*. 2025;271:116999. https://doi.org/10.1016/j.bios.2024.116999.
4 Hu L, Wang J, Xu Y, et al. A critical review of persimmon-derived pectin: innovations in extraction, structural characterization, biological potentials, and health-promoting effects. *Food Chem*. 2025;463(Pt 4):141453. https://doi.org/10.1016/j.foodchem.2024.141453
5 Proctor ML, Murphy PA. Herbal and dietary therapies for primary and secondary dysmenorrhoea. *Cochrane Database Syst Rev*. 2001;3:CD002124. https://doi.org/10.1002/14651858.CD002124.
6 Fatima K, Alokaili H, Alharbi M, et al. Effectiveness of fish oil in controlling inflammation in adult patients undergoing hemodialysis: a systematic review and meta-analysis. *SAGE Open Med*. 2024; 12: 20503121241275467. https://doi.org/10.1177/20503121241275467
7 Isman MB. Botanical insecticides, deterrents, and repellents in modern agriculture and an increasingly regulated world. *Annu Rev Entomol*. 2006;51:45-66. https://doi.org/10.1146/annurev.ento.51.110104.151146.
8 Wu Q, Zhou J. The application of polyphenols in food preservation. *Adv Food Nutr Res*. 2021;98:35-99. https://doi.org/10.1016/bs.afnr.2021.02.005.
9 Muteeb G, Rehman MT, Shahwan M, et al. Origin of antibiotics and antibiotic resistance, and their impacts on drug development: a narrative review. *Pharmaceuticals (Basel)*. 2023;16(11):1615. https://doi.org/10.3390/ph16111615.
10 Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J Nat Prod*. 2020;83(3):770-803. https://doi.org/10.1021/acs.jnatprod.9b01285.
11 Lyddiard D, Jones GL, Greatrex BW. Keeping it simple: lessons from the golden era of antibiotic discovery. *FEMS Microbiol Lett*. 2016;363(8):fnw084. https://doi.org/10.1093/femsle/fnw084.
12 Hautbergue T, Jamin EL, Coste F, et al. From genomics to metabolomics, moving toward an integrated strategy for the discovery of fungal secondary metabolites. *Nat Prod Rep*. 2018;35(2):147-173. https://doi.org/10.1039/c7np00032d.
13 Zhang S, Liu Y, Wang Q, et al. Cyclic natural product oligomers: diversity and (bio)synthesis of macrocycles. *Chem Soc Rev*. 2025;54(1):396-464. https://doi.org/10.1039/d2cs00909a.
14 Yang R, Chen H, Li X, et al. Factors influencing surimi gelling properties and natural additive-based gel fortification strategies: a review. *Compr Rev Food Sci Food Saf*. 2025;24(1):e70067. https://doi.org/10.1111/1541-4337.70067.
15 Sun P, Wang Y, Zhang L, et al. In pursuit of lead innovation: pharmaceutically important and distinct amide-free succinate dehydrogenase inhibitors. *J Med Chem*. 2025;68(2):1051-1067. https://doi.org/10.1021/acs.jmedchem.4c02757.
16 Shang Y, Zhao H. Research progress of Chinese medicinal monomers in the process of melanoma occurrence. *Pharm Biol*. 2025;63(1):53-67. https://doi.org/10.1080/13880209.2024.2445695.
17 Wang L, Liu R, Zhang Y, et al. Non-targeted metabonomics reveals the effect of linalyl alcohol on Brochothrix thermosphacta and its potential application. *Food Res Int*. 2025;201:115549. https://doi.org/10.1016/j.foodres.2024.115549.
18 Husain I, Ahmad A, Khan MS, et al. Evaluation of bioaccessibility, metabolic clearance and interaction with xenobiotic receptors (PXR and AhR) of cinnamaldehyde. *Food Chem (Oxf)*. 2025;10:100237. https://doi.org/10.1016/j.fochms.2024.100237.
19 Fu W, Yang J, Li C, et al. Copper-luteolin nanocomplexes for mediating multifaceted regulation of oxidative stress, intestinal barrier, and gut microbiota in inflammatory bowel disease. *Bioact Mater*. 2025;46:118-133. https://doi.org/10.1016/j.bioactmat.2024.12.004.
20 Liang X, Chen Y, Wang H, et al. Herbal micelles-loaded ROS-responsive hydrogel with immunomodulation and microenvironment reconstruction for diabetic wound healing. *Biomaterials*. 2025;317:123076. https://doi.org/10.1016/j.biomaterials.2024.123076.
21 Subhash A, Kumar S, Sharma M, et al. Date seeds polysaccharides as novel capping agents for selenium nanoparticles: synthesis, characterization, stability, biological activities, and gut microbiota modulation. *Food Chem*. 2025;470:142746. https://doi.org/10.1016/j.foodchem.2024.142746.
22 Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. 2015;349(6245):255-260. https://doi.org/10.1126/science.aaa8415.
23 Mjolsness E, DeCoste D. Machine learning for science: state of the art and future prospects. *Science*. 2001;293(5537):2051-2055. https://doi.org/10.1126/science.293.5537.2051.
24 Roy K, Gaur M, Sheth A, et al. Neurosymbolic artificial intelligence (why, what, and how). *IEEE Intell Syst*. 2023;38(3):56-62. https://doi.org/10.1109/MIS.2023.3268724.
25 Aliferis C, Simon GJ. Foundations and properties of AI/ML systems[M]// Simon GJ, Aliferis C. Artificial intelligence and machine learning in health care and medical sciences: best practices and pitfalls. *Cham (CH): Springer*, 2024: 33-94.https://doi.org/10.1007/978-3-031-39355-6_2
26 van Santen JA, Poynton EF, Iskakova D, et al. The natural products atlas: an open access knowledge base for microbial natural products discovery. *ACS Cent Sci*. 2019;5(11):1824-1833. https://doi.org/10.1021/acscentsci.9b00806.
27 Lanz J, Riedl R. Merging allosteric and active site binding motifs: *de novo* generation of target selectivity and potency *via* natural-product-derived fragments. *ChemMedChem*. 2015;10(3):451-454. https://doi.org/10.1002/cmdc.201402478.
28 Wassermann AM, Lounkine E, Hoepfner D, et al. A screening pattern recognition method finds new and divergent targets for drugs and natural products. *ACS Chem Biol*. 2014;9(7):1622-1631. https://doi.org/10.1021/cb5001839.
29 Ozturk H, Ozkirimli E, Ozgur A. Exploring chemical space using natural language processing methodologies for drug discovery. *Drug Discov Today*. 2020;25(4):689-705. https://doi.org/10.1016/j.drudis.2020.01.020.
30 Boaro A, Caglioti C, Mosconi L, et al. Structure-function-guided design of synthetic peptides with anti-infective activity derived from wasp venom. *Cell Rep Phys Sci*. 2023;4(7):101456. https://doi.org/10.1016/j.xcrp.2023.101459.
31 Torres MDT, Silva ON, Alves ESF, et al. Structure-function-guided exploration of the antimicrobial peptide polybia-CP identifies activity determinants and generates synthetic therapeutic candidates. *Commun Biol*. 2018;1:221. https://doi.org/10.1038/s42003-018-0224-2.

32 Pedron CN, Torres MDT, Silva ON, et al. Novel designed VmCT1 analogs with increased antimicrobial activity. *Eur J Med Chem*. 2017;126:456-463. https://doi.org/10.1016/j.ejmech.2016.11.040.
33 Torres MT, Silva ON, Alves ESF, et al. Antimicrobial activity of leucine-substituted decoralin analogs with lower hemolytic activity. *J Pept Sci*. 2017;23(11):818-823. https://doi.org/10.1002/psc.3029.
34 Baum ZJ, Yu X, Ayala PY, et al. Artificial intelligence in chemistry: current trends and future directions. *J Chem Inf Model*. 2021;61(7):3197-3212. https://doi.org/10.1021/acs.jcim.1c00619.
35 Stokes JM, Yang K, Swanson K, et al. A deep learning approach to antibiotic discovery. *Cell*. 2020;181(2):475-483. https://doi.org/10.1016/j.cell.2020.01.021.
36 Kretschmer F, Rastogi R, Korshunova M, et al. Coverage bias in small molecule machine learning. *Nat Commun*. 2025;16(1):554. https://doi.org/10.1038/s41467-024-55462-w.
37 Cai Z, Liu X, Wang Y, et al. Style mixup enhanced disentanglement learning for unsupervised domain adaptation in medical image segmentation. *Med Image Anal*. 2024;101:103440. https://doi.org/10.1016/j.media.2024.103440.
38 Hastings J, Owen G, Dekker A, et al. ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res*. 2016;44(D1):D1214-D1219. https://doi.org/10.1093/nar/gkv1031.
39 Wang Y, Xiao J, Suzek TO, et al. PubChem's BioAssay database. *Nucleic Acids Res*, 2012, 40(Database issue): D400-D412.https://doi.org/10.1093/nar/gkr1132
40 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30. https://doi.org/10.1093/nar/28.1.27.
41 van Santen JA, Poynton EF, Iskakova D, et al. The natural products atlas 2.0: a database of microbially-derived natural products. *Nucleic Acids Res*. 2022;50(D1):D1317-D1323. https://doi.org/10.1093/nar/gkab941.
42 Irwin JJ, Sterling T, Mysinger MM, et al. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model*. 2012;52(7):1757-1768. https://doi.org/10.1021/ci3001277.
43 Terlouw BR, Blin K, Navarro-Munoz JC, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res*. 2023;51(D1):D603-D610. https://doi.org/10.1093/nar/gkac1049.
44 Lee YY, Wang M, Liu X, et al. HypoRiPPAtlas as an atlas of hypothetical natural products for mass spectrometry database search. *Nat Commun*. 2023;14(1):4219. https://doi.org/10.1038/s41467-023-39905-4.
45 Wishart DS, Guo A, Oler E, et al. NP-MRD: the natural products magnetic resonance database. *Nucleic Acids Res*. 2022;50(D1):D665-D677. https://doi.org/10.1093/nar/gkab1052.
46 Gallo K, Goede A, Preissner R, et al. SuperNatural 3.0-a database of natural products and natural product-based derivatives. *Nucleic Acids Res*. 2023;51(D1):D654-D659. https://doi.org/10.1093/nar/gkac1008.
47 Winkelmann J, Schall P, Dijkstra M, et al. Corrected article: simulation and observation of line-slip structures in columnar structures of soft spheres [Phys. Rev. E 96, 012610 (2017)]. *Phys Rev E*, 2017, 97(5-2): 059902. https://doi.org/10.1103/physreve.97.059902
48 Niven DJ, Leigh JP, Stelfox HT. Les considerations ethiques relatives a la desadoption d'aspects inefficaces ou nuisibles des soins de sante. *Healthc Manage Forum*. 2016;29(5):218-222. https://doi.org/10.1177/0840470416666576.
49 Wang P, Li X, Zhang Y, et al. Gut microbiota-derived 4-hydroxyphenylacetic acid from resveratrol supplementation prevents obesity through SIRT1 signaling activation. *Gut Microbes*. 2025;17(1):2446391. https://doi.org/10.1080/19490976.2024.2446391.
50 Bhanukiran K, Priya V. Exploring multitarget potential of piper nigrum fruit constituents for Alzheimer's disease: an AI-driven strategy. *Phytomedicine*. 2024;135:156108. https://doi.org/10.1016/j.phymed.2024.156108.
51 Chihomvu P, Moyo M, Ntuli N, et al. Phytochemicals in drug discovery-a confluence of tradition and innovation. *Int J Mol Sci*. 2024;25(16):8945. https://doi.org/10.3390/ijms25168792.
52 Nicolaou KC, Hale CRH, Nilewski C, et al. Constructing molecular complexity and diversity: total synthesis of natural products of biological and medicinal importance. *Chem Soc Rev*. 2012;41(15):5185-5238. https://doi.org/10.1039/c2cs35116a.
53 Ory L, Nazih EH, Daoud L, et al. Targeting bioactive compounds in natural extracts-development of a comprehensive workflow combining chemical and biological data. *Anal Chim Acta*. 2019;1070:29-42. https://doi.org/10.1016/j.aca.2019.04.038.
54 Liu P, Wang Y, Zhang L, et al. Wutou decoction: a latest review on molecular mechanisms, clinical studies, quality control, pharmacokinetic studies, pharmacological effects, toxicity, and target prediction. *J Ethnopharmacol*. 2025;341:119307. https://doi.org/10.1016/j.jep.2024.119307.
55 Sun L, Chen H, Wang J, et al. Polysaccharides from traditional Chinese medicine and their nano-formulated delivery systems for cancer immunotherapy. *Carbohydr Polym*. 2025;357:123416. https://doi.org/10.1016/j.carbpol.2025.123416.
56 Saldivar-Gonzalez FI, Sanchez-Cruz N, Medina-Franco JL, et al. Natural product drug discovery in the artificial intelligence era. *Chem Sci*. 2022;13(6):1526-1546. https://doi.org/10.1039/D1SC04471K.
57 Meunier M, Schinkovitz A, Derbre S. Current and emerging tools and strategies for the identification of bioactive natural products in complex mixtures. *Nat Prod Rep*. 2024;41(11):1766-1786. https://doi.org/10.1039/d4np00006d.
58 Wolfender JL, Marti G, Thomas A, et al. Accelerating metabolite identification in natural product research: toward an ideal combination of liquid chromatography-high-resolution tandem mass spectrometry and NMR profiling, in silico databases, and chemometrics. *Anal Chem*. 2019;91(1):704-742. https://doi.org/10.1021/acs.analchem.8b05112.
59 Menna M, Imperatore C, D'Aniello F, et al. Challenges in the configuration assignment of natural products a case-selective perspective. *Nat Prod Rep*. 2019;36(3):476-489. https://doi.org/10.1039/c8np00053k.
60 Kussmann M, Abe Cunha DH, Berciano S. Bioactive compounds for human and planetary health. *Front Nutr*. 2023;10:1193848. https://doi.org/10.3389/fnut.2023.1193848.
61 Agrawal P, Khater S, Gupta M, et al. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res*. 2017;45(W1):W80-W88. https://doi.org/10.1093/nar/gkx408.
62 Gangwal A, Lavecchia A. AI-driven drug discovery for rare diseases. *J Chem Inf Model*. 2025;65(5):2214-2231. https://doi.org/10.1021/acs.jcim.4c01966.
63 Gangwal A, Lavecchia A. Artificial intelligence in natural product drug discovery: current applications and future perspectives. *J Med Chem*. 2025;68(4):3948-3969. https://doi.org/10.1021/acs.jmedchem.4c01257.
64 Howarth A, Ermanis K, Goodman JM. DP4-AI automated NMR data analysis: straight from spectrometer to structure. *Chem Sci*. 2020;11(17):4351-4359. https://doi.org/10.1039/d0sc00442a.
65 Duan FL, Wang Y, Zhang L, et al. AI-driven drug discovery from natural products. *Advanced Agrochem*. 2024;3(3):185-187. https://doi.org/10.1016/j.aac.2024.06.003.
66 Marcarino MO, Ciceri D, Sarotti AM. A critical review on the use of DP4 + in the structural elucidation of natural products: the good, the bad and the ugly.a practical guide. *Nat Prod Rep*. 2022;39(1):58-76. https://doi.org/10.1039/d1np00030f.
67 Marcarino MO, Zanardi MM, Sarotti AM. NMR calculations with quantum methods: development of new tools for structural elucidation and beyond. *Acc Chem Res*. 2020;53(9):1922-1932. https://doi.org/10.1021/acs.accounts.0c00365.
68 Ozcelik R, Chithrananda S, Voelker A, et al. Chemical language modeling with structured state space sequence models. *Nat Commun*. 2024;15(1):6176. https://doi.org/10.1038/s41467-024-50469-9.
69 Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discov Today*. 2015;20(3):318-331. https://doi.org/10.1016/j.drudis.2014.10.012.
70 Nguyen L, Wang Y, Zhang X, et al. iANP-EC: identifying anticancer natural products using ensemble learning incorporated with evolutionary computation. *J Chem Inf Model*, 2022, 62(21): 5080-5089.https://doi.org/10.1021/acs.jcim.1c00920
71 Astolfi R, Mencarelli A, Cangiotti M, et al. Essential oils as antimicrobials against acinetobacter baumannii: experimental and literature data to definite predictive quantitative composition-activity relationship models using machine learning algorithms. *J Chem Inf Model*. 2025;65(3):1378-1391. https://doi.org/10.1021/acs.jcim.4c02389.
72 Hasselgren C, Oprea TI. Artificial intelligence for drug discovery: are we there yet?. *Annu Rev Pharmacol Toxicol*. 2024;64:527-550. https://doi.org/10.1146/annurev-pharmtox-040323-040828.
73 Wang B, Zhang Y, Li X, et al. Explainable deep learning and virtual evolution identifies antimicrobial peptides with activity against multidrug-resistant human pathogens. *Nat Microbiol*. 2025;10(2):332-347. https://doi.org/10.1038/s41564-024-01907-3.
74 Zhu S, Liu Y, Wang Q, et al. Computational advances in biosynthetic gene cluster discovery and prediction. *Biotechnol Adv*. 2025;79:108532. https://doi.org/10.1016/j.biotechadv.2025.108532.
75 Askr H, El-Sappagh S, Abuhmed T, et al. Exploring the anticancer activities of Sulfur and magnesium oxide through integration of deep learning and fuzzy rough set analyses based on the features of Vidarabine alkaloid. *Sci Rep*. 2025;15(1):2224. https://doi.org/10.1038/s41598-024-82483-8.
76 Chen D, Gao K, Nguyen DD, et al. Algebraic graph-assisted bidirectional transformers for molecular property prediction. *Nat Commun*. 2021;12(1):3521. https://doi.org/10.1038/s41467-021-23720-w.
77 Chiu YC, Chen HI, Zhang T, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. *BMC Med Genomics*. 2019;12(Suppl 1):18. https://doi.org/10.1186/s12920-018-0460-9.
78 Wang Y, Zhang L, Chen H, et al. DeepDRK: a deep learning framework for drug repurposing through kernel-based multi-omics integration. *Brief Bioinform*. 2021;22(5):bbab087. https://doi.org/10.1093/bib/bbab048.
79 Wang Y, Zhang X, Liu P, et al. New drug discovery and development from natural products: advances and strategies. *Pharmacol Ther*. 2024;264:108752. https://doi.org/10.1016/j.pharmthera.2024.108752.
80 Xu H, Wang J, Li Y, et al. Celastrol suppresses colorectal cancer via covalent targeting peroxiredoxin 1. *Signal Transduct Target Ther*. 2023;8(1):51. https://doi.org/10.1038/s41392-022-01231-4.
81 Tang Y, Chen H, Wang L, et al. A review of botany, ethnomedicine, phytochemistry, pharmacology and toxicology of Sarcandra species. *Phytomedicine*. 2024;135:156008. https://doi.org/10.1016/j.phymed.2024.156008.
82 Singla RK, Dubey AK, Garg A, et al. Natural products derived from medicinal plants and microbes might act as a game-changer in breast cancer: a comprehensive review of preclinical and clinical studies. *Crit Rev Food Sci Nutr*. 2023;63(33):11880-11924. https://doi.org/10.1080/10408398.2022.2097196.
83 Halverson JD. Micronutrient deficiencies after gastric bypass for morbid obesity. *Am Surg*. 1986;52(11):594-598. https://doi.org/10.1007/s00394-021-02619-8.
84 Niu Q, Wang Y, Zhang L, et al. TCMFP: a novel herbal formula prediction method based on network target's score integrated with semi-supervised learning genetic algorithms. *Brief Bioinform*. 2023;24(3):bbac598. https://doi.org/10.1093/bib/bbad102.
85 Ke JP, Wang Y, Chen H, et al. Unraveling anti-aging mystery of green tea in C. elegans: chemical truth and multiple mechanisms. *Food Chem*, 2024,

460(Pt 1): 140510.https://doi.org/10.1016/j.foodchem.2024.140510
86 Ma Z, Liu P, Wang Y, et al. Comprehensive evaluation of the combined extracts of epimedii folium and ligustri lucidi fructus for PMOP in ovariectomized rats based on MLP-ANN methods. *J Ethnopharmacol*. 2021;268:113563. https://doi.org/10.1016/j.jep.2020.113563.
87 Marconato L, Palla G, Heumos L, et al. SpatialData: an open and universal data framework for spatial omics. *Nat Methods*. 2025;22(1):58-62. https://doi.org/10.1038/s41592-024-02212-x.
88 Colby SM, Hoyt DW, Nothias LF, et al. Introducing molecular hypernetworks for discovery in multidimensional metabolomics data. *J Proteome Res*. 2024;23(11):4789-4801. https://doi.org/10.1021/acs.jproteome.3c00634.
89 Durand M, Roux B, Courtois B, et al. Unlocking plant bioactive pathways: omics data harnessing and machine learning assisting. *Curr Opin Biotechnol*. 2024;87:103135. https://doi.org/10.1016/j.copbio.2024.103135.
90 Bai W, Zhang Y, Liu X, et al. Machine learning assists prediction of genes responsible for plant specialized metabolite biosynthesis by integrating multi-omics data. *BMC Genomics*. 2024;25(1):418. https://doi.org/10.1186/s12864-024-10258-6.
91 Atanasov AG, Zotchev SB, Dirsch VM, et al. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov*. 2021;20(3):200-216. https://doi.org/10.1038/s41573-020-00114-z.
92 Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the genomics era. *Nat Rev Drug Discov*. 2015;14(2):111-129. https://doi.org/10.1038/nrd4510.
93 Mao Y, Wang Y, Zhang L, et al. Emerging artificial intelligence-driven precision therapies in tumor drug resistance: recent advances, opportunities, and challenges. *Mol Cancer*. 2025;24(1):123. https://doi.org/10.1186/s12943-025-02321-x.
94 Lin A, Chen H, Wang J, et al. From chaos to order: optimizing fecal microbiota transplantation for enhanced immune checkpoint inhibitors efficacy. *Gut Microbes*. 2025;17(1):2452277. https://doi.org/10.1080/19490976.2025.2452277.
95 Shi H, Zhang X, Wang Y, et al. DRExplainer: quantifiable interpretability in drug response prediction with directed graph convolutional network. *Artif Intell Med*. 2025;163:103101. https://doi.org/10.1016/j.artmed.2025.103101.
96 Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip Rev Data Min Knowl Discov*. 2019;9(4):e1312. https://doi.org/10.1002/widm.1312.
97 Vamathevan J, Clark D, Czodrowski P, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov*. 2019;18(6):463-477. https://doi.org/10.1038/s41573-019-0024-5.
98 Ye M, Liu Y, Wang Q, et al. Network pharmacology, molecular docking integrated surface plasmon resonance technology reveals the mechanism of Toujie Quwen Granules against coronavirus disease 2019 pneumonia. *Phytomedicine*. 2021;85:153401. https://doi.org/10.1016/j.phymed.2020.153401.
99 Deng Q, Wang Y, Zhang L, et al. Based on network pharmacology, molecular docking and experimental verification to reveal the mechanism of Andrographis paniculata against solar dermatitis. *Phytomedicine*. 2024;135:156025. https://doi.org/10.1016/j.phymed.2024.156025.
100 Chen H, Engkvist O, Wang Y, et al. The rise of deep learning in drug discovery. *Drug Discov Today*. 2018;23(6):1241-1250. https://doi.org/10.1016/j.drudis.2018.01.039.
101 Ma S, Liu J, Li W, et al. Machine learning in TCM with natural products and molecules: current status and future perspectives. *Chin Med*. 2023;18(1):43. https://doi.org/10.1186/s13020-023-00741-9.
102 Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017;542(7639):115-118. https://doi.org/10.1038/nature21056.
103 Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*. 2018;34(13):i457-i466. https://doi.org/10.1093/bioinformatics/bty294.
104 Gao L, Zhong L, Feng T, et al. An AI-driven strategy for active compounds discovery and non-destructive quality control in traditional Chinese medicine: a case of Xuefu Zhuyu Oral Liquid. *Talanta*. 2025;287:127627. https://doi.org/10.1016/j.talanta.2025.127627.
105 Maciejewska-Turska M, Georgiev MI, Kai G, et al. Advances in bioinformatic methods for the acceleration of the drug discovery from nature. *Phytomedicine*. 2025;139:156518. https://doi.org/10.1016/j.phymed.2025.156518.
106 Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589. https://doi.org/10.1038/s41586-021-03819-2.
107 Wang Y, Zhang Q, Chen Y, et al. Antitumor effects of immunity-enhancing traditional Chinese medicine. *Biomed Pharmacother*. 2020;121:109570. https://doi.org/10.1016/j.biopha.2019.109570.
108 Liu Y, Fang C, Luo J, et al. Traditional Chinese medicine for cancer treatment. *Am J Chin Med*. 2024;52(3):583-604. https://doi.org/10.1142/S0192415X24500253.
109 Wang K, Chen Q, Shao Y, et al. Anticancer activities of TCM and their active components against tumor metastasis. *Biomed Pharmacother*. 2021;133:111044. https://doi.org/10.1016/j.biopha.2020.111044.
110 Nogales C, Mamdouh ZM, List M, et al. Network pharmacology: curing causal mechanisms instead of treating symptoms. *Trends Pharmacol Sci*. 2022;43(2):136-150. https://doi.org/10.1016/j.tips.2021.11.004.
111 Jager N. Bioinformatics workflows for clinical applications in precision oncology. *Semin Cancer Biol*. 2022;84:103-112. https://doi.org/10.1016/j.semcancer.2020.12.020.
112 Cichonska A, Ravikumar B, Rahman R. AI for targeted polypharmacology: The next frontier in drug discovery. *Curr Opin Struct Biol*. 2024;84:102771. https://doi.org/10.1016/j.sbi.2023.102771.
113 Xerxa E, Bajorath J. Data-oriented protein kinase drug discovery. *Eur J Med Chem*. 2024;271:116413. https://doi.org/10.1016/j.ejmech.2024.116413.
114 Noor F, Asif M, Ashfaq UA, et al. Machine learning for synergistic network pharmacology: a comprehensive overview. *Brief Bioinform*. 2023;24(3):bbad120. https://doi.org/10.1093/bib/bbad120.
115 Isigkeit L, Hormann T, Schallmayer E, et al. Automated design of multi-target ligands by generative deep learning. *Nat Commun*. 2024;15(1):7946. https://doi.org/10.1038/s41467-024-52060-8.
116 Alshabrmi FM, Aba Alkhayl FF, Rehman A. Novel drug discovery: advancing Alzheimer's therapy through machine learning and network pharmacology. *Eur J Pharmacol*. 2024;976:176661. https://doi.org/10.1016/j.ejphar.2024.176661.
117 Li Z, Qu N, Zhou J, et al. KinomeMETA: a web platform for kinome-wide polypharmacology profiling with meta-learning. *Nucleic Acids Res*. 2024;52(W1):W489-W497. https://doi.org/10.1093/nar/gkae380.
118 Liang H, Fang C, Qiu M. The multi-target mechanism of action of Selaginella doederleinii Hieron in the treatment of nasopharyngeal carcinoma: a network pharmacology and multi-omics analysis. *Sci Rep*. 2025;15(1):159. https://doi.org/10.1038/s41598-024-83921-3.
119 Tian J, Qin S, Han J, et al. A review of the ethnopharmacology, phytochemistry, pharmacology and toxicology of Fructus Gardeniae (Zhi-zi). *J Ethnopharmacol*. 2022;289:114984. https://doi.org/10.1016/j.jep.2022.114984.
120 Luo Y, Gao F, Chang R, et al. Metabolomics based comprehensive investigation of Gardeniae Fructus induced hepatotoxicity. *Food Chem Toxicol*. 2021;153:112250. https://doi.org/10.1016/j.fct.2021.112250.
121 Shao W, Xu H, Zeng K, et al. Advances in liver organoids: replicating hepatic complexity for toxicity assessment and disease modeling. *Stem Cell Res Ther*. 2025;16(1):27. https://doi.org/10.1186/s13287-025-04139-2.
122 D'Alessio F, Koopman G, Houard S, et al. Workshop report: experimental animal models for universal influenza vaccines. *Vaccine*. 2018;36(46):6895-6901. https://doi.org/10.1016/j.vaccine.2018.10.024.
123 Tabernilla A, Dos Santos Rodrigues B, Pieters A, et al. In vitro liver toxicity testing of chemicals: a pragmatic approach. *Int J Mol Sci*. 2021;22(9):5038. https://doi.org/10.3390/ijms22095038.
124 Fruhwein H, Paul NW. "Lost in translation?" Animal research in the era of precision medicine. *J Transl Med*, 2025, 23(1): 152. https://doi.org/10.1186/s12967-025-06084-3.
125 Lin H, Li Z, Zeng T, et al. The crucial involvement of gamma-Mangostin and CYP1B1 in the mechanism underlying the toxicity caused by cigarette smoke extract: in silico and in vitro insights. *Toxicology*. 2025;510:154016. https://doi.org/10.1016/j.tox.2024.154016.
126 Jia C, Li X, Hu S, et al. Advanced mass-spectra-based machine learning for predicting the toxicity of traditional Chinese medicines. *Anal Chem*. 2025;97(1):783-792. https://doi.org/10.1021/acs.analchem.4c05311.
127 Scheufen Tieghi R, Moreira-Filho JT, Martin HJ, et al. A novel machine learning model and a web portal for predicting the human skin sensitization effects of chemical agents. *Toxics*. 2024;12(11):803. https://doi.org/10.3390/toxics12110803.
128 Debad SJ, Aungst J, Carstens K, et al. State of the science on assessing developmental neurotoxicity using new approach methods. *ALTEX*. 2025;42(1):121-144. https://doi.org/10.14573/altex.2410231.
129 Kim JW, Kim HS, Kim HR, et al. Next generation risk assessment of biocides (PHMG-p and CMIT/MIT)-induced pulmonary fibrosis using adverse outcome pathway-based transcriptome analysis. *J Hazard Mater*. 2024;476:134986. https://doi.org/10.1016/j.jhazmat.2024.134986.
130 Franzen L, Olsson Lindvall M, Huhn M, et al. Mapping spatially resolved transcriptomes in human and mouse pulmonary fibrosis. *Nat Genet*. 2024;56(8):1725-1736. https://doi.org/10.1038/s41588-024-01819-2.
131 Zhao Y, Yin N, Yang R, et al. Recent advances in environmental toxicology: exploring gene editing, organ-on-a-chip, chimeric animals, and in silico models. *Food Chem Toxicol*. 2024;193:115022. https://doi.org/10.1016/j.fct.2024.115022.
132 Cao A, Zhang L, Bu Y, et al. Machine learning prediction of on/off target-driven clinical adverse events. *Pharm Res*. 2024;41(8):1649-1658. https://doi.org/10.1007/s11095-024-03742-x.
133 Yang Q, Fan L, Hao E, et al. Construction of an explanatory model for predicting hepatotoxicity: a case study of the potentially hepatotoxic components of Gardenia jasminoides. *Drug Chem Toxicol*. 2025;48(1):107-119. https://doi.org/10.1080/01480545.2024.2364905.
134 Cai H, Zhang H, Zhao D, et al. FP-GNN: a versatile deep learning architecture for enhanced molecular property prediction. *Brief Bioinform*. 2022;23(6):bbac408. https://doi.org/10.1093/bib/bbac408.
135 Feinberg EN, Joshi E, Pande VS, et al. Improvement in ADMET prediction with multitask deep featurization. *J Med Chem*. 2020;63(16):8835-8848. https://doi.org/10.1021/acs.jmedchem.9b02187.
136 Chopra H, Annu, Shin DK, et al. Revolutionizing clinical trials: the role of AI in accelerating medical breakthroughs. *Int J Surg*. 2023;109(12):4211-4220. https://doi.org/10.1097/JS9.0000000000000705.
137 Gholap AD, Uddin MJ, Faiyazuddin M, et al. Advances in artificial intelligence for drug delivery and development: a comprehensive review. *Comput Biol Med*. 2024;178:108702. https://doi.org/10.1016/j.compbiomed.2024.108702.
138 Zhu T, Liu B, Chen N, et al. Artificial intelligence-driven innovations in pharmaceutical development and drug delivery systems. *Curr Top Med Chem*. 2025; 25(25):2937-2951. https://doi.org/10.2174/0115680266373236250411060857.
139 Serrano DR, Luciano FC, Anaya BJ, et al. Artificial intelligence (AI) applications in drug discovery and drug delivery: revolutionizing personalized medicine. *Pharmaceutics*. 2024;16(10):1328. https://doi.org/10.3390/pharmaceutics16101328.

140 Liu Y, Ahmed Smait D, Yaseen Naser A, et al. Optimization and validation of drug solubility by development of advanced artificial intelligence models. *J Mol Liq*. 2023;372:121113. https://doi.org/10.1016/j.molliq.2022.121113.

141 MacIntyre MR, Cockerill RG, Mirza OF, et al. Ethical considerations for the use of artificial intelligence in medical decision-making capacity assessments. *Psychiatry Res*. 2023;328:115466. https://doi.org/10.1016/j.psychres.2023.115466.

142 Bittner MI, Farajnia S. AI in drug discovery: applications, opportunities, and challenges. *Patterns (N Y)*. 2022;3(6):100529. https://doi.org/10.1016/j.patter.2022.100529.

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

144 Kim S. Getting the most out of PubChem for virtual screening. *Expert Opin Drug Discov*. 2016;11(9):843-855. https://doi.org/10.1080/17460441.2016.1216967.

145 Chevrette MG, Gavrilidou A, Mantri S, et al. The confluence of big data and evolutionary genome mining for the discovery of natural products. *Nat Prod Rep*. 2021;38(11):2024-2040. https://doi.org/10.1039/d1np00013f.

146 Lynch C, Sakamuru S, Ooka M, et al. High-throughput screening to advance in vitro toxicology: accomplishments, challenges, and future directions. *Annu Rev Pharmacol Toxicol*. 2024;64:191-209. https://doi.org/10.1146/annurev-pharmtox-112122-104310.

147 Yanofsky C. Transcription attenuation: once viewed as a novel regulatory strategy. *J Bacteriol*. 2000;182(1):1-8. https://doi.org/10.1128/JB.182.1.1-8.2000.

148 Zhu H, Rusyn I, Richard A, et al. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. *Environ Health Perspect*. 2008;116(4):506-513. https://doi.org/10.1289/ehp.10573.

149 Li Q, Wang Y, Bryant SH. A novel method for mining highly imbalanced high-throughput screening data in PubChem. *Bioinformatics*. 2009;25(24):3310-3316. https://doi.org/10.1093/bioinformatics/btp589.

150 Cai C, Wang S, Xu Y, et al. Transfer learning for drug discovery. *J Med Chem*. 2020;63(16):8683-8694. https://doi.org/10.1021/acs.jmedchem.9b02147.

151 Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *npj Digit Med*. 2020;3:119. https://doi.org/10.1038/s41746-020-00323-1.

152 Kadurin A, Nikolenko S, Khrabrov K, et al. 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): 3098-3104. https://doi.org/10.1021/acs.molpharmaceut.7b00346.

153 Wong F, Omori S, Li A, et al. An explainable deep learning platform for molecular discovery. *Nat Protoc*. 2025; 20:1020–1056.https://doi.org/10.1038/s41596-024-01084-x.

154 Pereira TO, Abbasi M, Arrais JP. ABIET: an explainable transformer for identifying functional groups in biological active molecules. *Comput Biol Med*. 2025;187:109740. https://doi.org/10.1016/j.compbiomed.2025.109740.

155 Budhkar A, Song Q, Su J, et al. Demystifying the black box: a survey on explainable artificial intelligence (XAI) in bioinformatics. *Comput Struct Biotechnol J*. 2025;27:346-359. https://doi.org/10.1016/j.csbj.2024.12.027.

156 Cath C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philos Trans A Math Phys Eng Sci*. 2018;376(2133):20180080. https://doi.org/10.1098/rsta.2018.0080.

157 Chen P, Dong W, Wang J, et al. Interpretable clinical prediction *via* attention-based neural network. *BMC Med Inform Decis Mak*. 2020;20(Suppl 3):131. https://doi.org/10.1186/s12911-020-1110-7.

158 Tang Q, Nie F, Zhao Q, et al. A merged molecular representation deep learning method for blood-brain barrier permeability prediction. *Brief Bioinform*. 2022;23(5):bbac357. https://doi.org/10.1093/bib/bbac357.

159 Qadri YA, Shaikh S, Ahmad K, et al. Explainable artificial intelligence: a perspective on drug discovery. *Pharmaceutics*. 2025;17(9):1119. https://doi.org/10.3390/pharmaceutics17091119.

160 Zhang Z, Beck MW, Winkler DA, et al. Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. *Ann Transl Med*. 2018;6(11):216. https://doi.org/10.21037/atm.2018.05.32.

161 Yin Z, Zhang M, Liu R, et al. Explainable machine learning models enhance prediction of PFAS bioactivity using quantitative molecular surface analysis-derived representation. *Water Res*. 2025;280:123500. https://doi.org/10.1016/j.watres.2025.123500.

162 Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng*. 2018;2(10):749-760. https://doi.org/10.1038/s41551-018-0304-0.

163 Schneider P, Walters WP, Plowright AT, et al. Rethinking drug design in the artificial intelligence era. *Nat Rev Drug Discov*. 2020;19(5):353-364. https://doi.org/10.1038/s41573-019-0050-3.

164 Gaulton A, Hersey A, Nowotka M, et al. The ChEMBL database in 2017. *Nucleic Acids Res*. 2017;45(D1):D945-D954. https://doi.org/10.1093/nar/gkw1074.

165 Sakano K, Furui K, Ohue M. NPGPT: natural product-like compound generation with GPT-based chemical language models. *J Supercomput*. 2024;81(1):352. https://doi.org/10.1007/s11227-024-06860-w.

166 Huang K, Xiao C, Glass LM, et al. MolTrans: molecular interaction transformer for drug-target interaction prediction. *Bioinformatics*. 2021;37(6):830-836. https://doi.org/10.1093/bioinformatics/btaa880.

167 Shen X, Zeng T, Chen N, et al. NIMO: a natural product-inspired molecular generative model based on conditional transformer. *Molecules*. 2024;29(8):1867. https://doi.org/10.3390/molecules29081867.

168 Srinivasan S, Bajorath J. Protocol to generate dual-target compounds using a transformer chemical language model. *STAR Protoc*. 2025;6(1):103584. https://doi.org/10.1016/j.xpro.2024.103584.

169 Chen L, Li Y, Ma Y, et al. Multiscale graph equivariant diffusion model for 3D molecule design. *Sci Adv*. 2025;11(16):eadv0778. https://doi.org/10.1126/sciadv.adv0778.

170 Wang Y, Song M, Liu F, et al. Artificial intelligence using a latent diffusion model enables the generation of diverse and potent antimicrobial peptides. *Sci Adv*. 2025;11(6):eadp7171. https://doi.org/10.1126/sciadv.adp7171.

171 Jung W, Goo S, Hwang T, et al. Absorption distribution metabolism excretion and toxicity property prediction utilizing a pre-trained natural language processing model and its applications in early-stage drug development. *Pharmaceuticals (Basel)*. 2024;17(3):382. https://doi.org/10.3390/ph17030382.

172 Ancajas CMF, Oyedele AS, Butt CM, et al. Advances, opportunities, and challenges in methods for interrogating the structure activity relationships of natural products. *Nat Prod Rep*. 2024;41(10):1543-1578. https://doi.org/10.1039/d4np00009a.

173 Li Y, Duan Z, Li Z, et al. Data and AI-driven synthetic binding protein discovery. *Trends Pharmacol Sci*. 2025;46(2):132-144. https://doi.org/10.1016/j.tips.2024.12.002.

174 Gangwal A, Ansari A, Ahmad I, et al. Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities. *Front Pharmacol*. 2024;15:1331062. https://doi.org/10.3389/fphar.2024.1331062.

175 Guo SB, Meng Y, Lin L, et al. Artificial intelligence alphafold model for molecular biology and drug discovery: a machine-learning-driven informatics investigation. *Mol Cancer*. 2024;23(1):223. https://doi.org/10.1186/s12943-024-02140-6.

176 He F, Aebersold R, Baker MS, et al. pi-HuB: the proteomic navigator of the human body. *Nature*, 2024, 636(8042): 322-331. https://doi.org/10.1038/s41586-024-08280-5.

177 Zdouc MM, Blin K, Louwen NLL, et al. MIBiG 4.0: advancing biosynthetic gene cluster curation through global collaboration. *Nucleic Acids Res*. 2025; 53(D1):D678-D690. https://doi.org/10.1093/nar/gkae1115.

178 Mak KK, Pichika MR. Artificial intelligence in drug development: present status and future prospects. *Drug Discov Today*. 2019;24(3):773-780. https://doi.org/10.1016/j.drudis.2018.11.014.

179 Price WN, 2nd, Cohen IG. Privacy in the age of medical big data. *Nat Med*. 2019;25(1):37-43. https://doi.org/10.1038/s41591-018-0272-7.

180 Yaraghi N, Gopal RD. The role of HIPAA omnibus rules in reducing the frequency of medical data breaches: insights from an empirical study. *Milbank Q*. 2018;96(1):144-166. https://doi.org/10.1111/1468-0009.12314.

181 Appelbaum PS. Privacy in psychiatric treatment: threats and responses. *Am J Psychiatry*. 2002;159(11):1809-1818. https://doi.org/10.1176/appi.ajp.159.11.1809.

182 Floridi L, Cowls J, King TC, et al. How to design AI for social good: seven essential factors. *Sci Eng Ethics*. 2020;26(3):1771-1796. https://doi.org/10.1007/s11948-020-00213-5.

183 Paul D, Sanap G, Shenoy S, et al. Artificial intelligence in drug discovery and development. *Drug Discov Today*. 2021;26(1):80-93. https://doi.org/10.1016/j.drudis.2020.10.010.

184 Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24-29. https://doi.org/10.1038/s41591-018-0316-z.

185 Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health*. 2018;8(2):020303. https://doi.org/10.7189/jogh.08.020303.

186 Force AAT, Parasa S, Berzin T, et al. Consensus statements on the current landscape of artificial intelligence applications in endoscopy, addressing roadblocks, and advancing artificial intelligence in gastroenterology. *Gastrointest Endosc*. 2025;101(1):2-9. https://doi.org/10.1016/j.gie.2023.12.003.

187 Silver LL. Challenges of antibacterial discovery. *Clin Microbiol Rev*. 2011;24(1):71-109. https://doi.org/10.1128/CMR.00030-10.

188 Merwin NJ, Mousa WK, Dejong CA, et al. DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc Natl Acad Sci U S A*. 2020;117(1):371-380. https://doi.org/10.1073/pnas.1901493116.

189 Louwen JJR, van der Hooft JJJ. Comprehensive large-scale integrative analysis of omics data to accelerate specialized metabolite discovery. *mSystems*.2021;6(4):e0072621.https://doi.org/10.1128/mSystems.00726-21.

190 Huber F, van der Burg S, van der Hooft JJJ, et al. MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra. *J Cheminform*. 2021;13(1):84. https://doi.org/10.1186/s13321-021-00558-4.

191 Kim HW, Wang M, Leber CA, et al. NPClassifier: a deep neural network-based structural classification tool for natural products. *J Nat Prod*. 2021;84(11):2795-2807. https://doi.org/10.1021/acs.jnatprod.1c00399.

192 Schneider P, Schneider G. A computational method for unveiling the target promiscuity of pharmacologically active compounds. *Angew Chem Int Ed Engl*.2017;56(38):11520-11524.https://doi.org/10.1002/anie.201706376.

193 Sahayasheela VJ, Lankadasari MB, Dan VM, et al. Artificial intelligence in microbial natural product drug discovery: current and emerging role. *Nat Prod Rep*. 2022;39(12):2215-2230. https://doi.org/10.1039/d2np00035k.

194 Bender A, Cortes-Ciriano I. Artificial intelligence in drug discovery: what is realistic, what are illusions? Part 1: ways to make an impact, and why we are not there yet. *Drug Discov Today*. 2021;26(2):511-524. https://doi.org/10.1016/j.drudis.2020.12.009.

195 Cech NB, Medema MH, Clardy J. Benefiting from big data in natural products: importance of preserving foundational skills and prioritizing data quality. *Nat Prod Rep*. 2021;38(11):1947-1953. https://doi.org/10.1039/d1np00061f.

196 van Santen JA, Kautsar SA, Medema MH, et al. Microbial natural product databases: moving forward in the multi-omics era. *Nat Prod Rep*. 2021; 38(1):264-278. https://doi.org/10.1039/d0np00053a.

197 Fang Y, Zhang Q, Zhang N, et al. Knowledge graph-enhanced molecular contrastive learning with functional prompt. *Nat Mach Intell*. 2023;5(5):542-553. https://doi.org/10.1038/s42256-023-00654-0.

198 Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell*. 2020;2(10):573-584. https://doi.org/10.1038/s42256-020-00236-4.

199 Ding Y, Fei Y, Lu B. Emerging new concepts of degrader technologies. *Trends Pharmacol Sci*. 2020;41(7):464-474. https://doi.org/10.1016/j.tips.2020.04.005.

200 Rishikesh Magar, Yuyang Wang, Cooper Lorsung，et al. AugLiChem: data augmentation library of chemical structures for machine learning. *Mach Learn Sci Technol*. 2022;3:045015. https://doi.org/10.1088/2632-2153/ac9c84.

201 Huanbutta K, Burapapadh K, Kraisit P, et al. Artificial intelligence-driven pharmaceutical industry: a paradigm shift in drug discovery, formulation development, manufacturing, quality control, and post-market surveillance. *Eur J Pharm Sci*. 2024;203:106938. https://doi.org/10.1016/j.ejps.2024.106938.

202 Farghali H, Kutinova Canova N, Arora M. The potential applications of artificial intelligence in drug discovery and development. *Physiol Res*. 2021;70(Suppl4):S715-S722. https://doi.org/10.33549/physiolres.934765.

203 Sandeep Ganesh G, Kolusu AS, Prasad K, et al. Advancing health care *via* artificial intelligence: from concept to clinic. *Eur J Pharmacol*. 2022;934:175320. https://doi.org/10.1016/j.ejphar.2022.175320.

204 Macarron R, Banks MN, Bojanic D, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov*. 2011;10(3):188-195. https://doi.org/10.1038/nrd3368.

205 Graziano G, Stefanachi A, Contino M, et al. Multicomponent reaction-assisted drug discovery: a time- and cost-effective green approach speeding up identification and optimization of anticancer drugs. *Int J Mol Sci*. 2023;24(7):6581. https://doi.org/10.3390/ijms24076581.

206 Ganem B. Strategies for innovation in multicomponent reaction design. *Acc Chem Res*. 2009;42(3):463-472. https://doi.org/10.1021/ar800214s.

207 Dos Santos JA, de Castro PP, de Oliveira KT, et al. Multicomponent reactions applied to total synthesis of biologically active molecules: a short review. *Curr Top Med Chem*. 2023;23(11):990-1003. https://doi.org/10.2174/1568026623666230403102437.

208 Koromina M, Pandi MT, Patrinos GP. Rethinking drug repositioning and development with artificial intelligence, machine learning, and omics. *OMICS*. 2019;23(11):539-548. https://doi.org/10.1089/omi.2019.0151.

209 Zhang X, Han X, Xiang T, et al. From high resolution tandem mass spectrometry to pollutant toxicity AI-based prediction: a case study of 7 endocrine disruptors endpoints. *Environ Sci Technol*. 2025;59(9):4505-4517. https://doi.org/10.1021/acs.est.4c11417.

210 Kim J, Lee SJ, Jung D, et al. Development of a deep neural network model based on high throughput screening data for predicting synergistic estrogenic activity of binary mixtures for consumer products. *J Hazard Mater*. 2025;489:137650. https://doi.org/10.1016/j.jhazmat.2025.137650.

211 Hang R, Yao X, Bai L, et al. Evolving biomaterials design from trial and error to intelligent innovation. *Acta Biomater*. 2025;197:29-47. https://doi.org/10.1016/j.actbio.2025.03.013.

212 Xu Y, Ma S, Cui H, et al. AGILE platform: a deep learning powered approach to accelerate LNP development for mRNA delivery. *Nat Commun*. 2024;15(1):6305. https://doi.org/10.1038/s41467-024-50619-z.

213 Li F, Mou M, Li X, et al. DrugMAP 2.0: molecular atlas and pharma-information of all drugs. *Nucleic Acids Res*. 2025;53(D1):D1372-D1382. https://doi.org/10.1093/nar/gkae791.

214 Ferrarezi AA, de Souza JVP, Maigret B, et al. Rational design and synthesis of pyrazole derivatives as potential SARS-CoV-2 M(pro) inhibitors: an integrated approach merging combinatorial chemistry, molecular docking, and deep learning. *Bioorg Med Chem*. 2025;120:118095. https://doi.org/10.1016/j.bmc.2025.118095.

215 Basnet BB, Zhou ZY, Wei B, et al. Advances in AI-based strategies and tools to facilitate natural product and drug development. *Crit Rev Biotechnol*, 2025: 1-32. https://doi.org/10.1080/07388551.2025.2478094.

216 Vasighi M, Romanova J, Nedyalkova M. A multilevel approach for screening natural compounds as an antiviral agent for COVID-19. *Comput Biol Chem*. 2022;98:107694. https://doi.org/10.1016/j.compbiolchem.2022.107694.

217 Zhang H, Guo J, Li H, et al. Machine learning for artemisinin resistance in malaria treatment across *in vivo-in vitro* platforms. *iScience*. 2022;25(3):103910. https://doi.org/10.1016/j.isci.2022.103910.

218 Jia XN, Wang WJ, Yin B, et al. Deep learning promotes the screening of natural products with potential microtubule inhibition activity. *ACS Omega*. 2022;7(32):28334-28341. https://doi.org/10.1021/acsomega.2c02854.

219 Wang H, Xie M, Rizzi G, et al. Identification of sclareol as a natural neuroprotective ca(v) 1.3-antagonist using synthetic parkinson-mimetic gene circuits and computer-aided drug discovery. *Adv Sci (Weinh)*. 2022;9(7):e2102855. https://doi.org/10.1002/advs.202102855.

220 Yang R, Zhao G, Yan B. Discovery of novel c-jun n-terminal kinase 1 inhibitors from natural products: integrating artificial intelligence with structure-based virtual screening and biological evaluation. *Molecules*. 2022;27(19):6249. https://doi.org/10.3390/molecules27196249.

221 Barbosa H, Espinoza GZ, Amaral M, et al. Andrographolide: a diterpenoid from cymbopogon schoenanthus identified as a new hit compound against trypanosoma cruzi using machine learning and experimental approaches. *J Chem Inf Model*. 2024;64(7):2565-2576. https://doi.org/10.1021/acs.jcim.3c01410.

222 Wang Y, Xiao J, Suzek TO, et al. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res*, 2009, 37(Web Server issue): W623-W633. https://doi.org/10.1093/nar/gkp456.

223 Sterling T, Irwin JJ. ZINC 15--ligand discovery for everyone. *J Chem Inf Model*, 2015, 55(11): 2324-2337. https://doi.org/10.1021/acs.jcim.5b00559.

224 Ogata H, Goto S, Sato K, et al. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 1999;27(1):29-34. https://doi.org/10.1093/nar/27.1.29.

225 Sorokina M, Steinbeck C. Review on natural products databases: where to find data in 2020. *J Cheminform*. 2020;12(1):20. https://doi.org/10.1186/s13321-020-00424-9.

226 Al Sharie AH, El-Elimat T, Al Zu'bi YO, et al. Chemical space and diversity of seaweed metabolite database (SWMD): a cheminformatics study. *J Mol Graph Model*. 2020;100:107702. https://doi.org/10.1016/j.jmgm.2020.107702.

227 Cockroft NT, Cheng X, Fuchs JR. STarFish: a stacked ensemble target fishing approach and its application to natural products. *J Chem Inf Model*. 2019;59(11):4906-4920. https://doi.org/10.1021/acs.jcim.9b00489.

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