Acta Pharmaceutica Sinica B 2026;16(7):4103−4127

Chinese Pharmaceutical Association
Institute of Materia Medica, Chinese Academy of Medical Sciences

Acta Pharmaceutica Sinica B

www.elsevier.com/locate/apsb
www.sciencedirect.com

ORIGINAL ARTICLE

Revolutionizing drug discovery from natural products: The roles of artificial intelligence and multi-omics in accelerating innovation

Boyang Wang[a,†], Qingyuan Liu[a,†], Weibo Zhao[a,b,†], Tingyu Zhang[a], Dingfan Zhang[a], Chayanis Sutcharitchan[a], Shao Li[a,*]

[a]*Institute for TCM-X, Department of Automation, Tsinghua University, Beijing 100084, China*
[b]*Institute of Information on Traditional Chinese Medicine, China Academy of Chinese Medical Sciences, Beijing 100700, China*

Received 18 June 2025; received in revised form 5 September 2025; accepted 20 October 2025

KEY WORDS

Artificial intelligence;
Natural products;
Multi-omics;
Drug development;
Bioactive compound discovery;
Compound derivatization and optimization;
Traditional Chinese medicine;
Network pharmacology

Abstract Natural products and their derivatives have long been crucial in drug therapy, especially in traditional medicine. However, challenges in screening, isolation, characterization, and optimization have slowed their development in the pharmaceutical industry. Recent advancements in artificial intelligence (AI) and multi-omics technologies are revitalizing this field. AI offers powerful tools for understanding natural compounds, enhancing molecular representations, and supporting tasks such as binding prediction, drug repurposing, and retrosynthesis. Moreover, generative models are aiding in natural product optimization and the creation of pseudo-natural compounds. At the same time, multi-omics technologies, including genomics, transcriptomics, proteomics, and metabolomics, have enabled high-throughput studies of plant traits, synthesis, regulatory mechanisms, and quality control, providing valuable data for AI model development. These advancements help accelerate the discovery of new compounds with medicinal potential. Furthermore, in the field of traditional Chinese medicine research, which is largely based on natural plant sources, AI systems exemplified by UNIQ system, combining AI and multi-omics, have been instrumental in mechanistic studies and new drug development. This study comprehensively discusses the algorithms and applications of AI and multi-omics technologies in the drug development of natural compounds and plants, as well as summarizing relevant databases which might provide high-quality data for the future development of AI algorithms targeting natural products.

This article is part of special issue entitled Machine Learning in Drug Discovery.

*Corresponding author.
E-mail address: shaoli@mail.tsinghua.edu.cn (Shao Li).
†These authors made equal contributions to this work.

Peer review under the responsibility of Chinese Pharmaceutical Association and Institute of Materia Medica, Chinese Academy of Medical Sciences.

https://doi.org/10.1016/j.apsb.2025.12.030

1. Introduction

Natural products have been a cornerstone of medicinal practices for centuries, deeply embedded in folklore for the treatment of various ailments[1]. Over the years, plant-derived compounds have garnered recognition as a rich source of therapeutic agents, celebrated for their vast structural diversity and pharmacological potential. Historically, natural products and their analogs have contributed significantly to pharmacotherapy, particularly in the treatment of complex diseases such as cancer[2], infections and inflammation[3]. The landscape of natural product drug discovery is shifting as new strategies enable more efficient identification and optimization of bioactive compounds. Natural products and their structural derivatives continue to serve as key candidates for novel drug development, with ongoing efforts focused on enhancing their pharmacological properties. Despite their proven efficacy, challenges in drug discovery—ranging from issues with screening, isolation, characterization, and optimization—have led to a decline in their pursuit by the pharmaceutical industry since the 1990s. Recent technological advancements, including improved analytical tools, genome mining, and microbial culturing techniques, have breathed new life into the search for natural product-based drug leads[4]. The incorporation of modern technologies such as target-based screening and multi-disciplinary approaches is addressing the limitations that once hindered the exploration of these compounds[5]. Today, the focus is on discovering new molecular entities—referred to as "new entities of natural product drug discovery"—that promise to tackle emerging global health threats, such as antimicrobial resistance. As such, natural products remain a vital and exciting area of research, offering substantial opportunities for the development of next-generation therapies.

To clarify the interconnections among key concepts central to this review, natural products are primarily derived from natural plants, particularly medicinal plants, which produce bioactive substances composed of various natural compounds. Many of these medicinal plants are integral to traditional Chinese medicine (TCM), where they have been used for centuries to treat various ailments. The natural compounds extracted from these plants form the basis of natural products, which are key candidates for modern drug discovery. By studying these compounds, researchers can develop new, nature-inspired therapies, advancing the field of pharmaceutical development.

Artificial intelligence (AI) is transforming the landscape of biomedical research by enhancing both the speed and scope of scientific discovery. In drug development, AI technologies, particularly generative models and large language models, are revolutionizing the process from target identification to clinical trials. These AI-driven methodologies are improving the efficiency and effectiveness of the drug discovery pipeline, allowing for more precise predictions and the generation of new molecular entities[6,7]. In cancer research, AI's potential applications extend to early detection, subtype classification, and the identification of new therapeutic targets, all of which are pivotal for transforming cancer diagnosis and treatment[8]. These AI tools rely on big data to create more accurate models for diagnosis and prognosis, offering groundbreaking potential for precision medicine. Similarly, AI has made significant strides in computational pathology, where deep learning techniques are automating clinical diagnosis, predicting patient outcomes, and identifying new biomarkers from tissue images. This progress is accelerating the integration of AI into clinical practice, although challenges related to widespread adoption and technical barriers remain[9]. Moreover, AI's ability to generate synthetic data through deep generative models is addressing critical data challenges in biomedicine, such as privacy concerns, bias, and data scarcity. By augmenting real-world data, synthetic data can improve fairness, reduce bias, and simulate previously unseen scenarios, enhancing the accuracy and inclusivity of AI models[10].

Multi-omics approaches are reshaping the landscape of biomedical research, offering powerful tools to study complex biological systems across various domains. The Earth BioGenome Project is leveraging multi-omics to sequence and catalog the genomes of all eukaryotic life, providing critical insights for conservation efforts, biodiversity management, and climate change research[11]. In the biomedical field, multi-omics technologies such as transcriptomics, epigenomics, proteomics, and metabolomics are essential for understanding the interactions of engineered nanoparticles with biological systems. These technologies enable high-resolution studies of nanoparticle biodistribution, cellular responses, and subcellular interactions, facilitating the development of targeted therapeutic strategies[12]. In immunology, advances in single-cell multi-omics are providing unprecedented detail on the development of the human immune system, from prenatal stages to adult immunity. By integrating genomic, transcriptomic, and proteomic data at single-cell resolution, researchers can map the complex transitions and interactions that underpin immune system development, offering new perspectives on human disease[13]. Furthermore, the integration of single-cell multi-omics in cancer research is transforming our understanding of tumor immunology, cell lineage tracing, and cellular spatial information, enabling more precise diagnostics and personalized treatment approaches[14,15].

The rapid growth of high-throughput omics technologies and diverse biomedical data has created vast datasets that AI, particularly multimodal foundation models, can integrate to enhance molecular biology research and clinical applications[16]. By combining genomics, transcriptomics, epigenomics, proteomics, and metabolomics, these models can generate holistic maps of cells and tissues, enabling breakthroughs in cell-type recognition, biomarker discovery, and gene regulation[17]. In oncology, multimodal fusion approaches that integrate molecular, radiology, and clinical data are essential for capturing disease complexity and improving personalized medicine[18]. Additionally, the growing availability of data from biobanks, wearable sensors, and electronic health records opens new opportunities for AI in personalized medicine, digital health, and remote care, although challenges related to data integration, privacy, and modeling remain[19].

AI is revolutionizing drug discovery by enhancing the prediction of biological activity, optimizing molecular design, and enabling *de novo* drug development[20-22]. Machine learning (ML)

models rapidly analyze vast datasets, facilitating virtual screening and accelerating the drug development process. In natural product drug discovery, AI's potential is equally transformative. By combining computational omics with AI, researchers can efficiently explore nature's molecular diversity, predict bioactivity, and identify novel drug candidates. This synergy promises to expedite the discovery of innovative therapies from natural sources, significantly enhancing the drug development pipeline[23].

In this review, we systematically introduce innovative AI algorithms in natural compounds and natural plants and their applications, with a focus on researches mainly from the past five years. The focus is on illustrating how AI algorithms help address key issues such as virtual screening and property prediction of natural compounds, as well as quality control, traits studies and synthesis patterns of natural plants, thereby better facilitating the exploration of the medicinal and pharmaceutical potential of natural products. The included studies mainly involve the development and application of AI algorithms addressing key issues in natural compounds and natural plants, with a primary focus on scientific problems related to the discovery of the potential medicinal value of natural products. In addition, we incorporated multiple high-quality datasets of natural compounds and natural plants. For natural compounds, we primarily curated databases related to drug responses, physicochemical properties, and target information. For natural plants, we focused on databases containing multi-omics data and information on metabolic and biosynthetic pathways. Additionally, we compiled databases related to TCM, including information on the compound composition of herbs and their targets.

2. AI-driven innovations to natural compound-based drug discovery

2.1. Advances in AI models for compounds analysis and prediction

AI is increasingly recognized as a transformative tool for compound analysis and prediction, underpinning advances across chemistry, pharmacology, and biology. Fig. 1 provides an overview of various AI methodologies used in natural compound research, showcasing how image, graph, and sequential data are processed with different neural network architectures to facilitate drug discovery tasks such as virtual screening and molecular generation. Central to these applications is molecular representation, which can be broadly classified into four categories. First, one-dimensional molecular fingerprints, often employed in conjunction with encoder-based architectures such as autoencoders, facilitate feature compression and abstraction[24,25]. Second, two-dimensional formats—including molecular graphs and structural depictions—are typically processed by graph neural networks or convolutional models[26-28]. Third, three-dimensional topological representations capture spatial molecular configurations and are leveraged by 3D graph neural networks and spatially-aware architectures[29-31]. Finally, hybrid models integrate multiple representational modalities, often through attention mechanisms or multi-branch frameworks, to enhance predictive performance[32,33]. Collectively, these strategies enable richer and more accurate molecular understanding, supporting a wide array of downstream tasks.

AI also plays a pivotal role in molecular structure[34,35] and property prediction[36-40], enabling rapid and accurate estimation of physicochemical attributes, bioactivity, toxicity, and other functional characteristics through deep learning and transfer learning approaches. In drug−target interaction prediction, AI-driven models integrate molecular features with omics data to estimate binding affinities and prioritize candidate targets, offering improvements in efficiency over conventional docking methods[41-44]. Drug repurposing and virtual screening likewise benefit from deep learning pipelines capable of identifying novel therapeutic uses for existing compounds across expansive chemical libraries[45-48]. Generative models represent a rapidly advancing frontier, facilitating the *de novo* design of drug-like molecules tailored to specific properties or structural motifs, thereby expanding the accessible chemical space[49-53]. Furthermore, AI is increasingly applied in chemical reaction prediction and synthesis planning, particularly in retrosynthesis, where models infer viable synthetic pathways and propose intermediates based on curated knowledge or data-driven

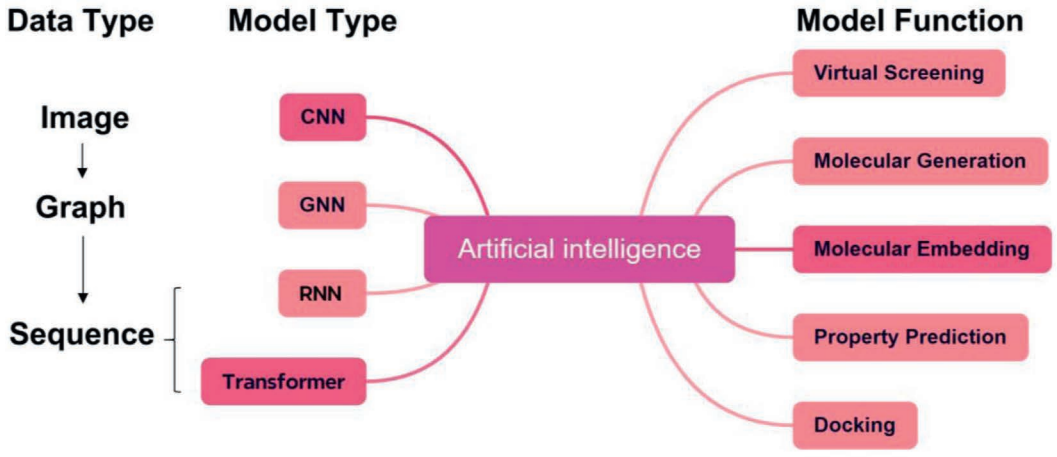

Figure 1 Overview of AI methodologies applied in natural compound research. This schematic illustrates the AI techniques based on different data representations: image data processed by Convolutional Neural Networks (CNN), graph data processed by Graph Neural Networks (GNN), and sequential data analyzed using Recurrent Neural Networks (RNN) and Transformers. These AI approaches facilitate essential drug discovery tasks such as virtual screening, molecular generation, molecular embedding, property prediction, and molecular docking, significantly advancing natural product-based drug innovation.

insights[54-56]. Finally, the integration of multi-modal and multi-omics data—encompassing genomic, transcriptomic, and other biological layers—with chemical information enhances compound analysis, enabling more accurate prediction of drug responses and compound efficacy in complex biological systems[57-59].

2.2. *Applications of AI in natural compounds drug discovery*

To systematically illustrate the role of AI in natural compound-based drug discovery, this review categorizes recent modeling approaches into four major areas: physics-based interaction and docking models, association and prediction models, generative models for novel compounds design, and models for natural compounds derivatization and optimization. Fig. 2 presents a comprehensive methodological framework that highlights these four interconnected domains, demonstrating how AI accelerates and innovates drug discovery from natural products.

2.2.1. *Physics-based interaction and docking models*

Computational strategies for natural compound-based drug discovery have advanced markedly through the integration of AI with physics-based interaction and docking models. These efforts primarily fall into two categories: molecular docking and binding affinity prediction. Both areas increasingly rely on traditional ML and deep learning techniques to improve predictive accuracy and computational efficiency. The primary challenge is accurately predicting the binding affinity between natural compounds and their targets, as traditional methods are computationally intensive and inefficient. Current solutions integrate machine learning techniques, significantly improving computational efficiency and predictive accuracy, thus accelerating virtual screening. However, the diversity and complexity of natural products pose challenges in capturing stereochemical features, which may lead to inaccurate predictions.

Molecular docking. In molecular docking, traditional ML approaches have proven effective in accelerating virtual screening. For example, V-Dock[60] integrates molecular property optimization with ML-based docking score prediction, using SMILES strings as input to bypass computationally intensive simulations while preserving high accuracy in binding conformation prediction. Another strategy[61] employs attention-based LSTM networks and models such as XGBoost to estimate docking scores, enabling efficient screening of millions of compounds with minimal training data. These approaches markedly reduce the computational demands of large-scale virtual screening. Building on this foundation, deep learning has further transformed molecular docking by facilitating more sophisticated analyses of three-dimensional structural information.

The GEM-Screen model[62] introduces a geometry-enhanced molecular representation coupled with active learning strategies, demonstrating high performance in identifying potential drug candidates. Advancing drug–target interaction modeling, the DTIGN framework[63] integrates graph neural networks with self-attention mechanisms, yielding improved bioactivity prediction. For pose classification, methods based on 3D convolutional neural networks and point cloud architectures[64] eliminate false-positive docking results. In parallel, GraphSite[65] leverages neural weighted message passing and graph-based representations to classify ligand binding sites. Furthermore, based on diffusion models, DIFFDOCK-PP[66] learns to transform unbound protein

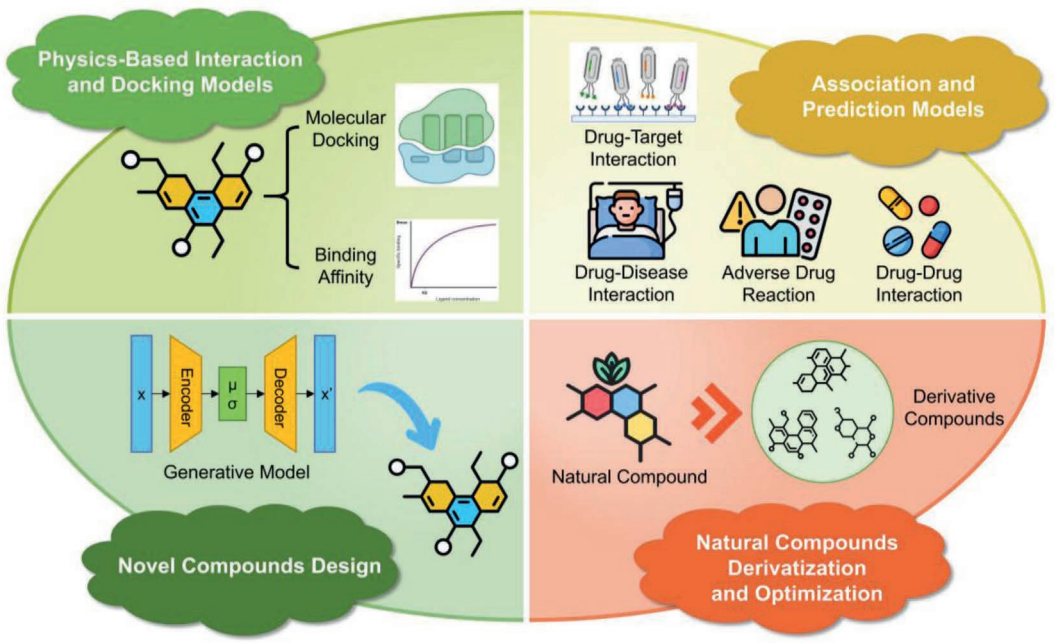

Figure 2 Methodological framework for AI-driven drug discovery from natural compounds. The diagram highlights four interconnected methodological domains: (1) physics-based interaction and docking models that employ molecular docking and binding affinity analysis; (2) association and prediction models utilized to explore drug-target interactions, drug-disease associations, adverse drug reactions, and drug-drug interactions; (3) novel compound design employing generative models such as variational autoencoders to facilitate the creation of new molecular structures; and (4) derivatization and optimization of natural compounds to generate derivatives with enhanced therapeutic potential. This comprehensive framework illustrates how AI-driven methodologies collectively accelerate and innovate drug discovery processes from natural products.

structures into their bound conformations through a reverse diffusion process. The model samples multiple docking poses and selects the most reliable one using a confidence model.

Binding affinity prediction. Binding affinity prediction has similarly benefited from recent advances in ML. Models leveraging ligand-residue interaction profiles and gradient boosting decision trees[67] have substantially outperformed traditional scoring functions, excelling in structure-based virtual screening by accurately predicting binding affinities and enhancing the identification of top ligands for drug discovery. In the realm of deep learning, SE-OnionNet[68] introduces a novel convolutional neural network that integrates squeeze-and-excitation modules to improve feature extraction from 3D protein−ligand complex structures. Similarly, the DeepAtom framework[69] utilizes 3D convolutional neural networks (CNNs) to capture atomic-level interaction patterns, advancing binding affinity predictions. PLANET[70] employs a graph neural network with a multi-task learning approach to predict multiple interaction properties. In the latest research progress, FlowDock[71] is a deep geometric generative model based on conditional flow matching, integrating protein−ligand docking and binding affinity prediction into a single framework. Specifically, it takes protein sequences and ligand SMILES as inputs, uses ESMFold to predict the unbound protein state and initializes ligands from a harmonic prior, maps unbound complexes to bound states *via* geometric flow matching, and outputs structural predictions, confidence scores, and binding affinities.

Collectively, these advancements signify a paradigm shift in computational drug discovery, enhancing both accuracy and efficiency while reducing the computational costs traditionally associated with physics-based simulations.

2.2.2. Association and prediction models

Association and prediction models are essential for identifying potential bioactive compounds in natural products and supporting drug repurposing by analyzing relationships between drugs, targets, and diseases. These models are significant in uncovering unknown drug−disease associations in natural product drug discovery. Current solutions utilize deep learning and graph neural networks to improve prediction accuracy. However, the generalization ability of models remains a limitation, particularly with limited data of specific categories.

Drug−target interaction prediction models. AI, particularly deep learning, has significantly advanced drug−target interaction (DTI) prediction, facilitating both drug discovery and repositioning. Recent innovations in this field can be categorized by their methodological advancements. These methods encompass a range of strategies, including attention-based models, graph-based techniques, contrastive learning, and ensemble-based approaches. Models like NFSA-DTI[72] and MINN-DTI[73] enhance feature representation through attention mechanisms. NFSA-DTI integrates neural fingerprints with self-attention to merge local and global molecular information, effectively identifying key interaction sites. MINN-DTI introduces mutual interaction modeling through an interacting-transformer module and communicative message passing neural network, offering a bidirectional representation of drug−target interactions. Further refinements are seen in approaches such as MCANet[74] and IGT[75], which improve attention-based frameworks. MCANet utilizes a lightweight MultiheadCrossAttention mechanism and addresses class imbalance and overfitting through PolyLoss. IGT applies an intermolecular graph transformer to model intricate spatial and topological data, enhancing the generalization and accuracy of binding predictions. Graph-based methods, including GCN-DTI[76], leverage graph convolutional networks to explicitly model drug−protein relationships, capturing broader structural context and improving predictive performance.

In contrastive learning, methods like CCL-ASPS[77] and SGCL-DTI[78] focus on optimizing sample representation and selection. CCL-ASPS combines collaborative contrastive learning with adaptive self-paced sampling to improve negative sample selection. SGCL-DTI employs supervised graph co-contrastive learning for heterogeneous networks, excelling in scenarios with limited data or cold-start problems. Ensemble-based approaches, such as EADTN[79] and MCANet-B[74], enhance robustness and prediction accuracy through model integration. EADTN reduces false negatives and improves interpretability *via* feature adaptation and clustering-enhanced fine-tuning, while MCANet-B combines multiple MCANet models, improving accuracy while maintaining computational efficiency. Together, these methods represent a spectrum of innovations that refine drug−target interaction predictions.

Drug−disease interaction prediction models. Beyond direct drug−target interactions, AI methodologies have also significantly enhanced predictions of drug−disease associations, a crucial area for effective drug repositioning and therapeutic discovery. Graph neural networks and knowledge graph-based methods are prominent for their high interpretability and predictive accuracy. The TxGNN model[80] integrates medical knowledge graphs with GNN and metric learning, enabling accurate zero-shot predictions for diseases lacking existing treatments. Similarly, DREAMwalk[81] leverages semantic multi-layer knowledge graphs through guilt-by-association random walks, improving predictive performance and interpretability by embedding drugs and diseases within a unified semantic space.

Multimodal and transcriptomic deep learning methods offer powerful approaches for utilizing comprehensive biological datasets. DLEPS[57] leverages transcriptional profiles to predict drug efficacy, providing insights into disease mechanisms and successfully validating predictions in experimental models. In parallel, STRGNN[82] integrates multimodal omics data, including transcriptomic, proteomic, and metabolomic layers, with graph neural networks and topological regularization, thus enhancing predictive accuracy by utilizing complementary data sources. Network topology models efficiently utilize existing biological network structures. One model[48] based on network target theory integrates extensive biological network topologies, by simulating drug action mechanisms, achieving high performance in predicting large-scale drug−disease interactions and synergistic drug combinations. The MTRD model[83] refines multi-scale heterogeneous network representations through random walk embedding and bi-directional LSTM modules, improving prediction capabilities. DDAGDL[84] employs geometric deep learning on heterogeneous information networks, capturing the complex non-Euclidean data structures and delivering robust predictive outcomes.

Neural collaborative filtering and convolutional neural network (CNN)-based methods offer powerful and efficient approaches to integrate and analyze diverse biomedical data for drug−disease prediction. DRWBNCF[85] applies a weighted bilinear neural collaborative filtering technique, integrating localized neighborhood information *via* graph convolutions to predict drug−disease associations effectively. CNN-based models, such as those leveraging molecular structures and clinical symptom data[86], provide straightforward, scalable methods for large-scale virtual screening. The IDDI-DNN[87] model integrates diverse similarity

matrices through CNNs, integrating multi-source drug, disease, and drug−disease association data. Specifically, it first constructs and calculates similarities for relevant matrices, uses similarity network fusion to integrate drug and disease similarity matrices, merges these with the association matrix into a single matrix, and then trains a convolutional neural network on this matrix to predict novel drug−disease associations, with dataset class balance achieved *via* SMOTE.

Adverse drug reaction prediction models. Accurate prediction of adverse drug reactions (ADRs) is vital for drug safety evaluation, patient safety, and successful drug development. Recent AI-driven advancements in ADR prediction can be broadly categorized into three methodological groups: heterogeneous graph and knowledge graph models, molecular structure-based models, and multi-modal data fusion techniques. Heterogeneous graph and knowledge graph representation models effectively utilize complex, interconnected biomedical data for comprehensive ADR prediction. PreciseADR[88] constructs heterogeneous patient-level graphs integrating drugs, diseases, and ADRs, enabling precise individual-specific ADR predictions through graph neural networks. Similarly, GCRS[89] employs heterogeneous graph convolutional autoencoders with attention mechanisms to fuse multiple specific and common graph topologies, effectively modeling diverse drug-side effect relationships. Besides, MVDSA[90] integrates multiple knowledge graphs and multi-view feature learning, using semantic encoding and attention mechanisms to capture nuanced relationships between drugs and ADRs comprehensively. Furthermore, TCSD[91] leverages heterogeneous graph transformers combined with capsule networks, effectively encoding attributes from diverse node types and semantic connections, substantially enhancing ADR prediction performance.

Molecular structure and deep chemical language models underscore detailed chemical information encoding for accurate ADR predictions. ADENet[92] integrates chemical substructure information with known drug−ADE networks, effectively bridging the gap to predict ADRs for new compounds beyond established interactions. MoLFormer-XL[93] utilizes deep chemical language modeling techniques applied to molecular SMILES strings, enabling precise identification of specific ADR types. iADRGSE[94] integrates molecular graph embedding with self-attention mechanisms, efficiently capturing structural sequence and graph information, particularly valuable at early drug development stages. Additionally, DSGAT[95] applies graph attention networks directly to drug molecular graphs, effectively addressing ADR frequency prediction challenges, especially for cold-start drugs with limited prior data.

Multi-modal data fusion and frequency prediction models specialize in integrating diverse data modalities to enhance frequency-based ADR categorization. MSSF[96] combines multi-source similarity data with a self-attention mechanism, improving accuracy in ADR frequency predictions while minimizing overfitting. Similarly, dual representation learning[97] merges protein target data, chemical similarity, and molecular graphs into dual representation vectors, refining frequency predictions, notably for previously unseen drugs. DeepSide[98] leverages multimodal deep learning frameworks that integrate chemical structures and gene expression profiles, effectively utilizing extensive biomedical datasets to achieve precise ADR frequency classifications. These approaches improve ADR prediction by addressing challenges such as data integration, structural complexity, and prediction accuracy, ultimately enhancing the identification of potential adverse effects in drug development.

Drug−drug interaction and combination prediction models. Accurate prediction of drug−drug interactions (DDIs) and drug combination synergies is pivotal for enhancing therapeutic efficacy and patient safety. Recent computational approaches for predicting DDIs and drug synergy can be categorized into four main methodological frameworks. Multi-task and multi-modal fusion models simultaneously optimize multiple predictive tasks while integrating diverse biomedical data modalities. DEML[99] leverages an ensemble-based multi-task neural network, combining chemical and transcriptomic data to simultaneously predict drug synergy, synergy classification, and adverse DDIs, effectively mitigating adverse interactions during combination therapy. Similarly, DeepTraSynergy[100] employs multimodal input from drug−target interactions, protein interactions, and cell-line data, utilizing transformer-based architectures to jointly predict drug−target relationships, toxicity, and synergistic effects, thereby enhancing predictive accuracy.

Graph and network propagation models exploit topological structures within biological networks to model complex drug interactions. GraphSynergy[101] utilizes graph convolutional networks on protein−protein interaction networks to explicitly encode high-order topological relationships, significantly improving the accuracy of anticancer drug synergy predictions. KGANSynergy[102] integrates knowledge graphs with attention mechanisms to capture hierarchical propagation of neighbor information, effectively identifying synergistic drug combinations. Additionally, network-propagation-inspired ML methods[103] leverage drug association networks to predict synergies across multiple cancer types, systematically validated through experimental and literature support. Furthermore, graph neural network-based methods, such as GNN-DDI[104], aggregate heterogeneous drug interaction attributes, successfully predicting adverse drug−drug interaction events. Biological pathway and gene expression-based mechanism models offer interpretable predictions grounded in underlying biological mechanisms. SynPathy[105] identifies pathway-level features associated with drug−target interactions, employing deep learning to predict drug synergies with explicit pathway proximity analyses. PathSynergy[106] integrates drug features, signaling pathways, and drug−target interactions *via* graph neural networks to predict liver cancer-specific drug synergies, validated through US Food and Drug Administration-approved drug combination predictions. Similarly, DeSIDE-DDI[107] harnesses drug-induced gene expression signatures, utilizing a gating mechanism to dynamically model drug interactions, achieving robust interpretability and predictive performance.

Substructure and 3D molecular structure models emphasize precise chemical-level interactions to refine DDI predictions. SSI-DDI[108] directly exploits substructure-substructure interactions *via* deep learning frameworks, effectively enhancing predictive capability through fine-grained chemical interaction modeling. The 3DGT-DDI[109] model incorporates 3D molecular graph structures and textual attention to predict DDIs. Specifically, it uses a 3D graph neural network to capture the spatial features of drug molecules and a pre-trained SCIBERT model for extracting textual data. The model combines these features through a hidden-layer fusion method, improving its prediction accuracy.

Collectively, recent AI-driven advances across various association and prediction tasks have significantly reshaped computational drug discovery. By leveraging sophisticated graph neural networks, attention mechanisms, deep chemical language models, and integrative multimodal frameworks, these approaches have effectively addressed complex biomedical problems including

drug−target, drug−disease, adverse drug reactions, and drug−drug interactions. These methodological innovations not only enhance predictive accuracy but also greatly improve interpretability, allowing for a more precise and systematic exploration of therapeutic candidates and safety profiles.

2.2.3. *Generative models for novel compounds design*

Generative models represent cutting-edge computational tools for exploring vast chemical spaces and enabling efficient *de novo* drug design. They offer an innovative approach for drug design by exploring chemical space and generating novel drug molecules. The significance of these models lies in providing diverse drug candidates, expanding the chemical space of natural products. The main challenge is balancing biological activity with synthetic feasibility when designing new molecules. Recent advances can be categorized into four prominent methodologies, each leveraging distinct computational strategies.

Graph representation and reinforcement learning models explicitly exploit molecular graph structures combined with reinforcement or evolutionary learning frameworks. For instance, reinforcement learning with graph-based generative models[110] guides pretrained models toward generating chemically diverse molecules with specific property profiles, even when these properties were not represented in the initial training data. Fragment-based approaches[111] further constrain molecular generation through graph fragments and evolutionary learning, effectively balancing multiple pharmacological objectives. Additionally, molecular substructure tree generative model[49] leverages variational autoencoders combined with graph-based architectures, enhancing molecular optimization and chemical validity.

Structure-guided and protein−ligand interaction-constrained models integrate explicit structural constraints and protein−ligand interaction patterns into molecular generation. Methods such as DeepLigBuilder[112] use 3D deep generative models to design molecules directly within protein binding sites, achieving chemically valid and target-specific molecular conformations. Similarly, docking-guided conditional generation models[113] employ pharmacophoric and docking score constraints to enhance structural novelty and biological activity. Interaction fingerprint-based generative models[114] further enforce ligand−protein interaction patterns, maintaining binding-mode consistency during *de novo* molecule generation.

Transformer and attention-based generative models have demonstrated remarkable advances in molecule generation by utilizing self-attention mechanisms to effectively process sequential and structural molecular representations. Transformer-based generative frameworks, such as TransAntivirus[115], allow efficient editing at the functional group level, enabling rapid antiviral drug analogue generation. Similarly, transformer graph variational autoencoders (TGVAE)[116] integrate transformers with graph neural networks and variational autoencoders to model complex molecular graph structures. Attention-augmented CycleGAN models[117] combine CycleGAN[118] architectures with embedded LSTM and attention mechanisms, improving chemical similarity between generated and reference molecules by effectively modeling long-term dependencies in molecular sequences.

Diffusion models and latent space-controlled generative models provide advanced mechanisms for molecular generation with precise structural and property control. Information-theoretically controlled latent spaces[119], enabled by gradient-based regularized search in variational autoencoders, rigorously reconstruct frameworks for generating property-specific molecules. Guided diffusion-based models, such as SILVR[120], iteratively refine latent representations to produce molecules fitting precise spatial constraints without extensive model retraining. Additionally, DrugDiff[121] is a small molecule diffusion model that generates molecules with specific properties by guiding the generation process. This process is directed by pre-trained property predictors, allowing for flexible optimization of molecular properties without the need to retrain the model. Geometry-based constrained autoencoders (GEOM-CVAE)[122] explicitly incorporate 3D structural constraints, enhancing molecular generation tailored to specific protein contexts. Specifically, the model introduces a geometric-based approach, embedding drug molecules'3D structures into images and protein targets into geometric graphs. It employs a constrained variational autoencoder to generate molecules by learning from these representations. The model consists of four key modules: a 3D structure-based molecular visual representation, a geometry-based protein graph representation, a constrained VAE, and a parser network to output SMILES strings.

Overall, these sophisticated generative models substantially advance our capabilities in novel molecular design, efficiently guiding the exploration of chemical space toward optimized drug-like candidates.

2.2.4. *Models for natural compounds derivatization and optimization*

Derivatization and optimization of natural compounds are critical steps in drug development, particularly when the initial activity of a natural product is insufficient. The challenge lies in modifying molecular structures while maintaining bioactivity. Recently, deep learning-based computational methods have emerged as powerful tools to overcome these barriers, significantly enhancing the exploration of novel derivatives and the optimization of natural product-based drug candidates. Several deep learning-driven approaches have been specifically developed to facilitate the derivatization and structural optimization of natural products. The DerivaPredict[123] platform integrates chemical and metabolic transformation rules with deep learning techniques, enabling the automated generation of novel natural product derivatives. The NIMO[124] model employs conditional transformer architectures with tailored motif-extraction methods, effectively capturing the structural complexity of natural products and enabling scaffold-based lead optimization with high structural diversity and biological relevance. The NPGPT model[125] harnesses a GPT-based chemical language model fine-tuned on natural product datasets, effectively producing structurally diverse and highly similar natural product-like molecules. Additionally, a large-scale generative model based on recurrent neural networks (RNN-LSTM)[126] produced over 67 million natural products-like molecules, greatly expanding the natural product chemical space and demonstrating strong validity and pharmaceutical potential through extensive cheminformatics validation.

Collectively, these advanced computational methods offer robust and efficient solutions for the derivatization and optimization of natural products, significantly enriching the chemical space exploration and accelerating the discovery of promising natural products-based drug candidates.

AI methodologies are playing a key role in overcoming the challenges of natural compound drug discovery. To provide a clearer comparison of these approaches, Table 1 summarizes the typical inputs/outputs, strengths, limitations, and applications of key AI models in this field.

Table 1 Comparison of AI methodologies in natural compound drug discovery.

Method category	Typical models	Major functions	Typical inputs	Typical outputs	Training data provenance	Strengths	Limitations/ weaknesses
Physics-based interaction & docking models	V-Dock, GEM-screen, DIFFDOCK-PP, DeepAtom, FlowDock	Molecular docking, binding affinity prediction	Molecular graphs, SMILES, 3D protein−ligand complex data	Docking scores, binding affinity predictions, optimized molecular conformations	Large protein-ligand interaction datasets, chemical libraries	Improved docking speed & accuracy; can model 3D protein −ligand interactions; diffusion/flow models capture binding poses better	Struggles with highly flexible/ complex natural products; still computationally demanding; limited generalizability to unseen scaffolds
Association & prediction models	NFSA-DTI, TxGNN, PreciseADR, GraphSynergy	Drug-target interaction prediction, drug−disease interaction prediction, adverse drug reaction prediction, drug−drug interaction and combination prediction	Graph representations, drug−target−disease interaction data, multi-omics data	Drug-target interaction scores, disease−drug association predictions, adverse drug reaction predictions, drug synergistic scores	Biological interaction networks, clinical datasets, omics datasets	Strong at integrating multi-omics & network data; high predictive accuracy	Dependent on high-quality labeled data; risk of overfitting; poor extrapolation in rare diseases or novel compounds
Generative models for novel compounds design	DeepLigBuilder, TransAntivirus, DrugDiff, molecular substructure tree generative model	*De novo* molecule generation, scaffold optimization	Molecular graphs, SMILES strings, chemical property data	Novel molecules with targeted properties, optimized molecular scaffolds	Chemical libraries, synthetic chemistry datasets	Explore vast chemical space; generate molecules with targeted properties; diffusion-based models enable flexible optimization	Balancing novelty *vs.* synthesizability remains difficult; May generate non-druggable compounds; requires careful property predictor guidance
Derivatization & optimization models	DerivaPredict, NIMO, NPGPT	Structural modification & optimization of NP scaffolds	Chemical transformation rules, molecular graphs, SMILES	Optimized molecules, natural product derivatives	Natural product datasets, synthetic chemistry datasets	Capture NP structural motifs; expand NP chemical space; enhance lead optimization efficiency	Limited by availability of transformation rules; May miss rare/complex NP modifications; biological relevance not always guaranteed

2.3. AI-driven natural compounds drug discovery

The practical integration of AI into the discovery of bioactive natural products represents a significant advancement in medicinal chemistry. Here, we summarize selected case studies that illustrate the capabilities and versatility of AI approaches in facilitating the identification and optimization of natural-compounds-inspired therapeutic agents (as shown in Table 2[127-130]). In a pioneering study, Merk et al.[127] developed an AI-driven approach utilizing recurrent neural networks to design novel compounds inspired by natural products targeting retinoid X receptors (RXR). The neural network was fine-tuned with bioactive natural-product templates, generating structures predicted to have RXR agonist activity. Synthesis and biological evaluation confirmed that two of the designed molecules exhibited notable RXR modulatory potency *in vitro*, thus demonstrating the potential of generative AI to efficiently discover novel, biologically active, natural-product-inspired compounds.

Another study[128] employed a multi-stage virtual screening framework integrating ML, molecular docking, and molecular dynamics simulations to identify natural product inhibitors of p38α MAP kinase. Initial screening used ML models trained on

Table 2 Comparison of AI-driven approaches for natural compounds -inspired drug discovery.

Method	Input	Training Data	Output	Validation	Limitation	Ref.
Recurrent neural network	SMILES strings of natural products; RXR agonist templates	541,555 bioactive small molecules from ChEMBL	RXR agonist candidates	*In vitro* pharmacological characterization, hybrid reporter gene assays, synthesis and evaluation of *de novo* designed compounds	Limited template structures	127
Machine learning, molecular docking and molecular dynamics simulation	MACCS molecular fingerprints	ChEMBL p38α MAPK inhibitors (2282 actives, 1340 inactives)	p38α MAPK inhibitors	Molecular docking, ADMET prediction, molecular dynamics simulation	Overfitting on training data, high false-positive rate	128
Self-attentive message passing neural network	Molecular graphs	Anti-osteoclastogenesis data (334 actives and 159 inactives), log*P* dataset with 21,364 compounds	Anti-osteoporosis natural product candidates	*In vitro* bioactivity testing, real-time qPCR, cytotoxicity assays, ADME profiling	Overfitting on training data, limited dataset	129
Rule-based generative model and target prediction	Chemical structure	25,563 commercial building blocks, 58 reaction schemes	COX-1 inhibitors	X-ray crystallography, COX inhibition assays, reporter gene assays, cytotoxicity assays, metabololipidomics	Synthetic accessibility, model generality	130

known inhibitors to filter a large library of natural products. Molecular docking and molecular dynamics simulations further narrowed down promising candidates. Two natural products (ZINC4260400 and ZINC8300300) showed strong interactions and favorable ADMET profiles, demonstrating stable binding to p38α MAP kinase *in silico*, indicating high potential for further experimental validation and drug development. In a related approach, Liu et al.[129] developed a deep learning model utilizing a self-attention-enhanced message-passing neural network to identify natural products with anti-osteoporosis activity. After computational screening, the team identified several potent compounds from natural product libraries. Subsequent *in vitro* assays demonstrated significant suppression of osteoclastogenesis and related gene expression, verifying the computational predictions. This case effectively showcased AI's capability in accelerating the discovery of bioactive natural compounds for complex diseases like osteoporosis.

Marine natural products have long been recognized as a valuable source of bioactive compounds with therapeutic potential, offering new avenues for drug discovery. Their approach is similar to those used in plant natural product drug discovery. For example, in marine natural product drug discovery, Lukas Friedrich et al.[130] proposed the DOGS algorithm to design synthetic molecules based on Marinopyrrole A. They also introduced the SPiDER algorithm to predict potential biological targets of these molecules, aiding in the identification of COX-1 inhibitors.

These studies highlight the transformative power of AI methodologies in discovering therapeutically valuable natural-product-inspired agents. Such approaches promise substantial efficiency gains in identifying bioactive compounds with favorable pharmaceutical profiles, underscoring AI's growing importance in modern drug discovery paradigms.

3. AI advances in natural plant-derived drug discovery

Natural plants, as an important component of medicinal plants, are also the primary source of natural medicinal compounds. In the development of drugs from natural plants, the metabolic synthesis mechanisms of the plant's active ingredients and quality control are critical issues. With the development of AI and multi-omics technologies, research on natural plants has been greatly aided, accelerating progress in areas such as plant traits, metabolic synthesis, and quality control. Fig. 3 illustrates the hierarchical application of multi-omics technologies—from genomics and transcriptomics to proteomics and metabolomics—providing a comprehensive approach to understanding the biological mechanisms of natural plants. Building on this, Fig. 4 presents an integrated framework that combines AI and multi-omics to systematically explore natural plant resources, addressing key aspects such as quality control, trait characterization, and the identification of medicinal properties.

3.1. Multi-omics in traits research of natural plants

The formation of complex traits in plants results from their genomes and activities at multiple molecular levels. Therefore, it is highly meaningful to link various plant omics data and these molecular activities to complex traits from a multi-omics perspective. Wang et al.[131] improved trait prediction in Arabidopsis by integrating genomic, transcriptomic, and methylomic data, discovering new flowering genes validated experimentally. Gene effects varied across genotypes, and multi-omics models outperformed single-omics ones by revealing complex gene interactions. This demonstrates the power of multi-omics for understanding complex traits. In higher plants, the complexity of systems and the components within and between species are rapidly dissected by omics technologies. Multi-omics datasets are integrated to infer and achieve a comprehensive understanding of the life processes of target organisms[132,133]. Besides, secondary metabolism in plants produces a vast diversity of small-molecule natural products. The discovery of operon-like gene clusters in plants offers new insights into the evolution of specialized metabolism and presents opportunities to rapidly advance the metabolic engineering of natural product biosynthesis[134].

By integrating advanced multi-omics approaches—transcriptomics, proteomics, metabolomics, and genomics—with microbial interactions, plant tolerance to abiotic stresses is

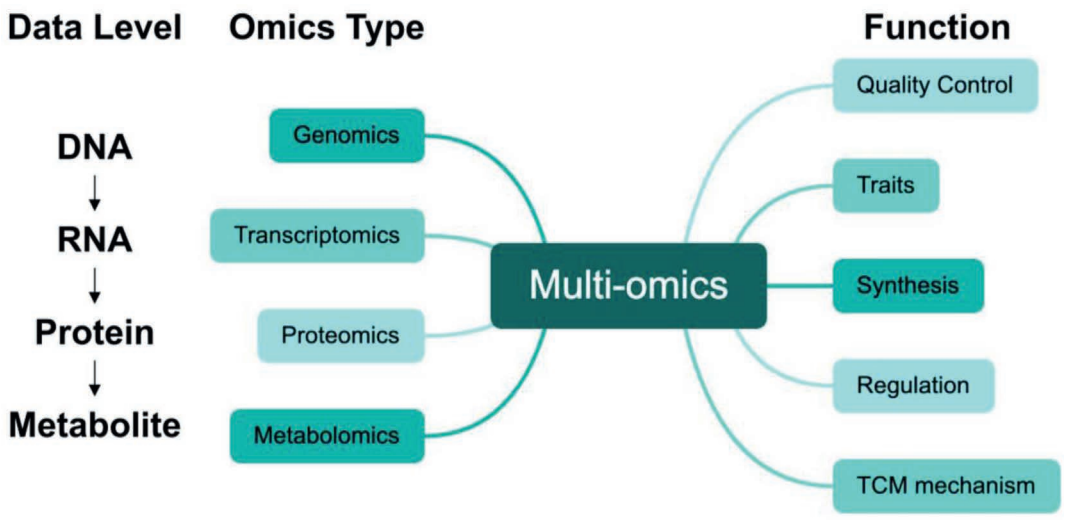

Figure 3 Overview of multi-omics methodologies applied in natural plant research. This schematic illustrates the hierarchical application of omics technologies—from DNA (genomics), RNA (transcriptomics), proteins (proteomics), to metabolites (metabolomics)—to study natural plants comprehensively. Multi-omics approaches support essential research areas such as quality control, trait identification, biosynthesis pathway elucidation, regulatory mechanism understanding, and exploration of traditional Chinese medicine (TCM) mechanisms, enabling holistic insights into plant-based drug discovery and optimization.

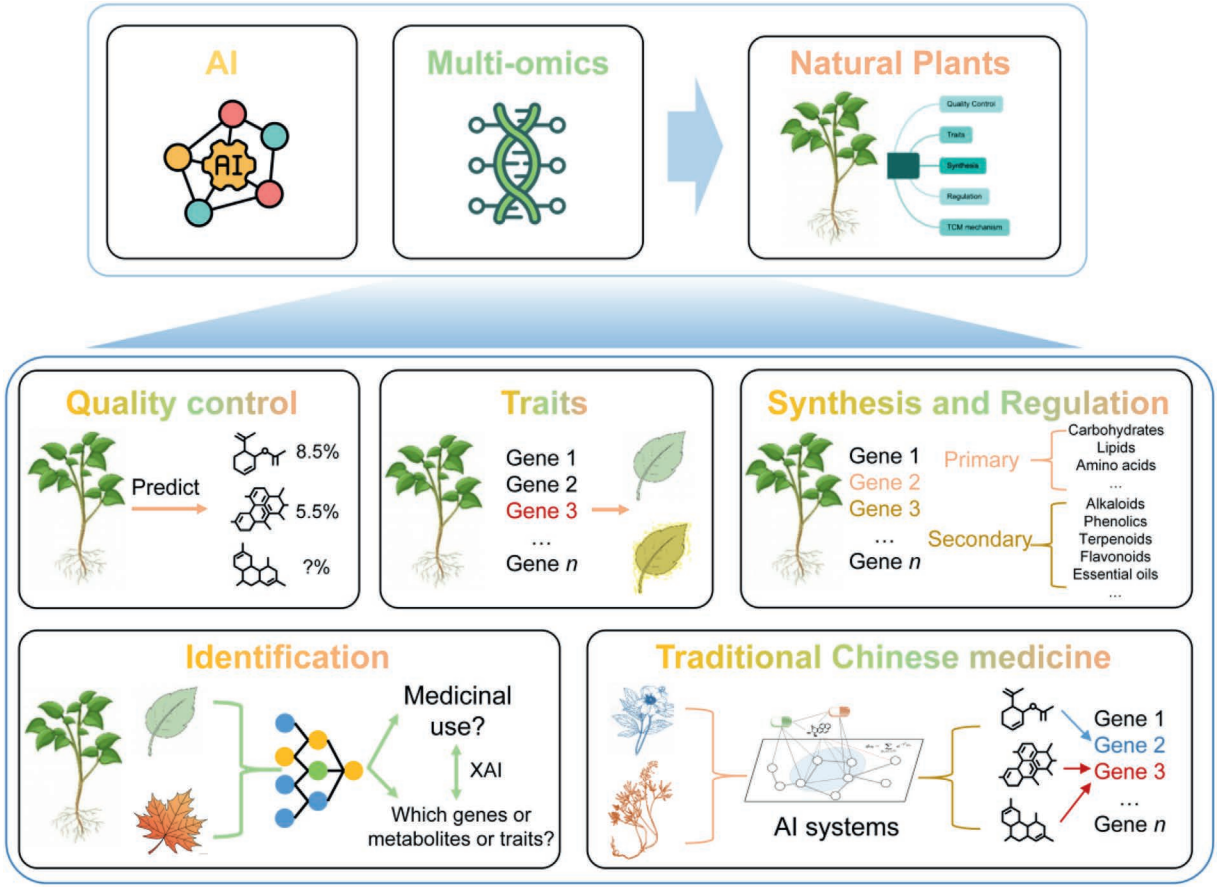

Figure 4 Integrated methodological framework combining AI and multi-omics for drug discovery from natural plants. The illustration demonstrates the synergistic application of AI and multi-omics analyses to systematically explore natural plant resources. This comprehensive framework addresses critical aspects including quality control through predictive modeling, trait characterization based on genetic markers, synthesis and regulatory pathways identification, medicinal identification, and uncovering mechanisms of traditional Chinese medicine. This integration accelerates the identification, optimization, and application of plant-derived pharmaceuticals.

significantly improved[135]. Genomics reveals genetic links between plants and beneficial microbes, aiding gene and pathway identification for stress tolerance. Transcriptomics compares gene expression in stressed *vs*. non-stressed plants to show microbial influence on stress responses. Proteomics identifies proteins involved in stress pathways and microbe-protein interactions supporting tolerance. Metabolomics analyzes small molecules to detect stress-induced metabolic changes and microbial metabolite roles. Multi-omics characterization of environmental microbes uses high-throughput technologies to reveal microbial composition, functions, and behaviors[136]. Advanced data analysis like network modeling and ML enables extraction of key insights and understanding complex interactions, expanding plant genetics and biology knowledge[137,138]. These methods accelerate microbial environment studies beyond traditional culturing, enabling exploration of genetic diversity and novel functions like toxin degradation. Multi-omics assesses DNA, RNA, proteins, and metabolites to evaluate environments affected by human activity. Transcriptomics detects anthropogenic gene expression changes; metaproteomics tracks related protein alterations[139]. Metagenomics profiles natural microbial communities and affected proteins. Metabolomics reveals stress-related metabolites and pathways supporting crop survival under heat stress[140], also assessing proteinaceous metabolites under normal conditions. Overall, multi-omics offers comprehensive insight into microbial community functions, surpassing single-omics limitations[141].

Plant metabolic gene clusters can be defined as containing two or more non-homologous and tightly linked genes that encode enzymes from the same biosynthetic pathway[142]. Furthermore, genes within the cluster are often coordinately regulated[143]. These characteristics make plant metabolic gene clusters valuable tools for functional identification of their associated biosynthetic pathways[144]. Meanwhile, the development of multi-omics approaches (combining two or more of genomics, transcriptomics, metabolomics, or epigenomics) has provided new strategies and opportunities for discovering natural product pathways. For example, a metabolite-based genome-wide association study (mGWAS) successfully identified a subspecies-specific diterpene (5,10-diketocasbene) gene cluster and a hydroxycinnamoyl-tyramine gene cluster in rice (*Oryza sativa* L.)[145,146]. Besides, by combining metabolomics with RNA sequencing analysis, a pathogen-responsive gene cluster responsible for the synthesis of falcarindiol was recently identified in tomato (*Solanum lycopersicum*)[147].

The application of multi-omics technologies in plant research offers significant potential for the development of natural plant-based medicines. By integrating genomics, transcriptomics, metabolomics, and proteomics data, we can gain a deeper understanding of key biological processes such as plant traits, stress tolerance, and metabolic pathways of natural product synthesis. These technologies not only help discover new medicinal compounds and metabolic pathways from plants but also accelerate the screening and functional validation of natural products. Through multi-omics analysis, researchers can more accurately predict the pharmacological effects of plants and efficiently identify compounds with medicinal potential. As the exploration of plant natural products deepens and multi-omics technologies continue to evolve, the development of natural plant medicines will become more efficient, driving the discovery and development of new drugs.

3.2. *Multi-omics enhancing quality control of natural plants*

The quality of natural plants plays a vital role in ensuring its efficacy and safety, directly affecting every stage of pharmaceutical production[148]. Key indicators for evaluating natural plants quality include its visual features and active ingredient content, both of which can be influenced by numerous internal and external factors (as shown in Table 3[149-157]). From the choice of germplasm sources and meticulous field management to harvesting, processing methods, and the environmental conditions during storage and transportation, each step contributes to the final quality of natural plants. This review provides a detailed examination of these factors affecting natural plants quality and explores how metabolomics technology is applied to monitor and control them, ultimately improving the quality of natural plants.

The origin and growth conditions of natural plants critically influence their quality, efficacy, and safety, which are paramount for ensuring authentic medicinal value[158]. Variations in environmental factors and cultivation methods cause differences in growth rate, yield, and bioactive compounds, leading to distinct regional characteristics in natural products. For example, *Tetrastigma hemsleyanum* polysaccharides exhibit antibacterial effects by interfering with bacterial glycolysis pathways, demonstrating potential for novel pharmacological assessment methods in traditional herbal medicines[159]. Comparative metabolomic and morphological analyses of goji berries from different Chinese climatic regions reveal species-specific chemical profiles rather than clear regional superiority, emphasizing the importance of chemical trait-based usage[151]. Ultra-performance liquid chromatography coupled with mass spectrometry and chemometric analysis enables precise authentication of medicinal species. This approach identified species-specific steroidal alkaloid markers in Fritillaria products, which can be used to ensure product authenticity and secure supply chains[149]. Metabolomic studies on purple

Table 3 Applications of multi-omics/AI technologies in quality control.

QC Question	Recommended omics/AI approach	Key metrics	Example study
Species authentication	UPLC−MS/High-throughput metabolomics; chemometric analysis/Targeted screening	Species-specific markers	149, 150
Geo-origin assessment	Comparative metabolomic analysis; morphological analysis	Metabolite profiling, region-specific chemical markers	151-154
Processing influence evaluation	UPLC−QTOF-MS/Metabolomic analysis; multivariate statistical analysis	Differentiation of crude/processed products; marker compounds; extraction yield; conversion efficiency	155, 156
Contamination detection	Metabolomic analysis; heavy metal profiling	Disruption of metabolic pathways; contaminant levels	157

sweet potato reveal that heavy metals such as uranium and cadmium disrupt plant metabolic pathways and hormone signaling, highlighting potential food safety risks from environmental contaminants[157]. Geographical differences in metabolite profiles and bioactivities have been observed in several medicinal plants. *Codonopsis lanceolata* grown in different regions showed significant variation in primary and secondary metabolites as well as antioxidant capacity, indicating distinct pharmacological potentials[152]. Similarly, dandelions from multiple Chinese provinces exhibited noticeable differences in phenolic compounds, with enriched biosynthesis pathways varying by origin[153]. A high-throughput metabolomics approach combined with targeted screening successfully differentiated four Pulsatilla Adans herbs at the chemical level, identifying key triterpenoid saponins as markers for species discrimination[150]. In *Eucommia ulmoides*, distinct metabolite profiles, antioxidant, and antibacterial activities among seven varieties were characterized, informing optimal utilization based on plant part and region of origin[154].

The processing of natural plants inevitably induces changes in their metabolite profiles, which can significantly affect their efficacy and safety[160]. Traditional processing methods—including baking, roasting, moistening, and burning—are designed to reduce toxicity, enhance therapeutic effects, facilitate preparation or storage, and purify the herbal drugs. For instance, in Astragali Radix, widely used in TCM, ultra-high performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry (UPLC−QTOF-MS) and multivariate statistical analysis revealed distinct chemical changes after different processing techniques. These changes enabled clear differentiation between crude and processed products and identified 15 significant marker compounds associated with the processing methods, providing a scientific basis to understand variations in pharmacological activity[155]. Similarly, Korean ginseng (*Panax ginseng* Meyer) undergoes chemical transformations during drying, steaming, and puffing. Puffing notably enhanced extraction yields and altered ginsenoside profiles by converting major ginsenosides into bioactive minor ones more rapidly than steaming, suggesting puffing as an efficient and economical processing technique to optimize ginseng's biological activity[156].

The quality of natural plants directly impacts their efficacy and safety, which in turn affects every stage of pharmaceutical production. Various factors, including the choice of germplasm sources, field management, harvesting, processing methods, and environmental conditions during storage and transportation, all play significant roles in determining plant quality. Multi-omics technologies, particularly metabolomics, provide essential tools for monitoring and controlling these factors, ultimately improving the quality of natural plants. For example, metabolomic analyses can accurately identify chemical traits of plants from different regions and processed by different methods, revealing differences in their pharmacological activity. This not only helps ensure product authenticity but also optimizes the development and utilization of plant-based medicines. For traditional herbal medicines, metabolomics aids in identifying key biomarkers, enabling the efficient development of plants with medicinal value and providing scientific foundations for drug quality control.

3.3. AI-based research of natural plants

Technological advances have revolutionized plant genotype and phenotype measurement, producing vast, complex datasets that demand sophisticated analytical tools. ML, a key AI component, has rapidly become essential in plant science, enabling multi-level phenotype analysis—from biochemical traits to yield—and bridging genotypes with phenotypes[161]. Leveraging AI enhances crop productivity by improving data collection, unlocking genetic diversity in genebanks, and advancing precision breeding through improved gene-editing accuracy and gene variant effect prediction, thus optimizing breeding programs for future environmental challenges and sustainability[162,163]. For instance, large-scale genomic studies in bread wheat have mapped numerous marker-trait associations across diverse environments, generating valuable genotype-phenotype resources to boost yield and stress resilience globally[164]. At the regulatory level, ML models using putative *cis*-regulatory elements outperform transcription factor-based models in predicting cell-type-specific stress responses in Arabidopsis, demonstrating broad applicability[165]. In genomic annotation, ML applied to epigenetic and omics data enhances gene expression classification in maize by linking methylation and histone modification patterns, adapting to line-specific variations[166]. For variant detection, deep learning tools like Clair3 combine pileup and full-alignment methods to deliver superior accuracy and speed in SNP calling from long reads, especially at low coverage[167]. Integrating remote sensing and ML, hyperspectral data analyzed by multilayer neural networks accurately predict wheat grain yield and identify key genetic loci[168]. Furthermore, deep learning models surpass traditional methods such as GBLUP in genomic prediction across multiple traits and environments, particularly with moderately large datasets, underscoring their growing value in practical breeding[169].

AI is revolutionizing phytochemical research by enabling efficient analysis of high-dimensional omics data, facilitating discovery and structural elucidation of novel plant metabolites, and advancing metabolomics and genomics with promising applications in medicine and agriculture[170]. In TCM, the integration of AI—especially ML—with hyperspectral imaging (HSI) has transformed quality control through enhanced data preprocessing, dimensionality reduction, and improved model accuracy, enabling faster and more precise evaluations while highlighting future challenges[171]. Meanwhile, ML-driven analysis of hyperspectral data is advancing photosynthesis research to boost sustainable crop yields[172] and enabling rapid, accurate food quality assessment through optimal wavelength selection, with deep learning showing promise for real-time applications despite existing hurdles[173]. Nutritional profiling benefits from near-infrared reflectance spectroscopy (NIRS) combined with modified partial least squares regression to deliver precise, high-throughput predictions for multiple biochemical traits in crops like pearl millet[174]. Chemical imaging, merging spectroscopy and spatial data, increasingly supports pharmaceutical quality control across product testing and industrial processes, although challenges like sample geometry and acquisition time remain[175]. Deep learning applications in agricultural hyperspectral image analysis exploit both spatial and spectral features for diverse tasks including ripeness estimation, classification, and disease detection, while addressing current limitations and outlining future directions[176]. Lastly, combining density functional theory with molecular dynamics and experimental NMR data enables precise conformational analysis of bioactive compounds, offering critical insights for theoretical and biological interaction studies[177].

Besides, facing the urgent challenges of climate change and a global population expected to reach 10 billion by 2050, accelerating varietal development is essential. Non-invasive high-throughput plant phenotyping (HTP), powered by advanced

imaging and deep learning, offers a transformative approach by abstracting complex sensor data and enhancing both throughput and accuracy. Recent progress in deep learning for HTP has identified optimal models, training methods, and public datasets, overcoming key obstacles and enabling multi-trait phenotyping and real-world breeding applications[178]. For example, multi-modal, multi-view, and time-lapse imaging systems allow non-destructive quantification of plant traits, addressing challenges like occlusion and lighting variation through robust analytical frameworks[179]. Hyperspectral imaging combined with CNN-based feature selection excels in accurately classifying crop seeds, as shown in wheat kernel variety identification[180]. Deep learning models—including multilayer perceptrons and CNNs—surpass traditional statistics in predicting complex genomic traits in crops like wheat[181]. UAV-mounted hyperspectral sensors paired with ML facilitate precise leaf chlorophyll estimation across scales and growth stages in maize and wheat[182]. In plant health, CNNs with data augmentation accurately detect and assess biotic stress in coffee leaves, enabling timely management[183]. Sparse 1D CNNs optimized by Bayesian techniques offer efficient genome-wide trait prediction, outperforming classical models[184]. Multi-class detection frameworks like Faster R-CNN enhance fruit condition recognition in dense apple canopies, improving robotic harvesting precision[185]. Real-time plant species identification powered by EfficientNet and mobile crowdsourcing achieves high accuracy in natural environments[186]. Lastly, novel deep learning tools such as DeepSignal-plant enable comprehensive genome-wide cytosine methylation profiling from Nanopore data, advancing plant epigenetics[187].

AI technologies, particularly ML, are rapidly advancing the development of natural plant-based medicines in plant science. Through large-scale genomic studies, plant phenotype analysis, and multi-level precision breeding, AI accelerates the improvement of plant traits, optimizing the yield and stress resilience of medicinal plants. The efficient analysis of multi-omics data facilitates the discovery and structural elucidation of new plant metabolites, driving the application of metabolomics and genomics in medicine and agriculture. The integration of ML with hyperspectral imaging (HSI) has enhanced quality control in TCM, enabling faster and more precise evaluations of plant material quality. Additionally, AI, through remote sensing and deep learning, assists in predicting crop yields, detecting plant health conditions, and improving the accuracy and efficiency of non-invasive high-throughput plant phenotyping. The integration of these technologies not only provides innovative tools for screening, identifying, and quality controlling natural plant medicines but also offers new approaches and methods for the development of plant-based drugs.

3.4. AI and multi-omics empowering therapeutics development from medicinal natural plants

3.4.1. Multi-omics in mechanism research of medicinal plants

Medicinal plants have been used as crucial sources of food and medicine throughout human history, with the World Health Organization (WHO) recognizing 21,000 species of medicinal value among approximately 300,000 plant species worldwide[188]. The advent of multi-omics technologies—encompassing genomics, transcriptomics, proteomics, and metabolomics—has significantly advanced our understanding of the complex genetic and biochemical networks that govern bioactive compound biosynthesis and regulation in medicinal plants. These integrated approaches enable the identification of gene pathways, regulatory sequences, and secondary metabolite biosynthesis, facilitating novel drug discovery and insights into environmental adaptation[189]. Fig. 5 illustrates the AI and multi-omics approaches applied to medicinal plant research. Although demand for herbal medicines is rising, challenges remain due to limited large-scale production and incomplete exploration of phytochemical diversity influenced by species variation, growth period, and environmental factors. Single-omics approaches often fail to capture the full complexity of biological regulation; thus, integrated multi-omics analyses have become increasingly important for elucidating molecular mechanisms and phenotypic plasticity shaped by

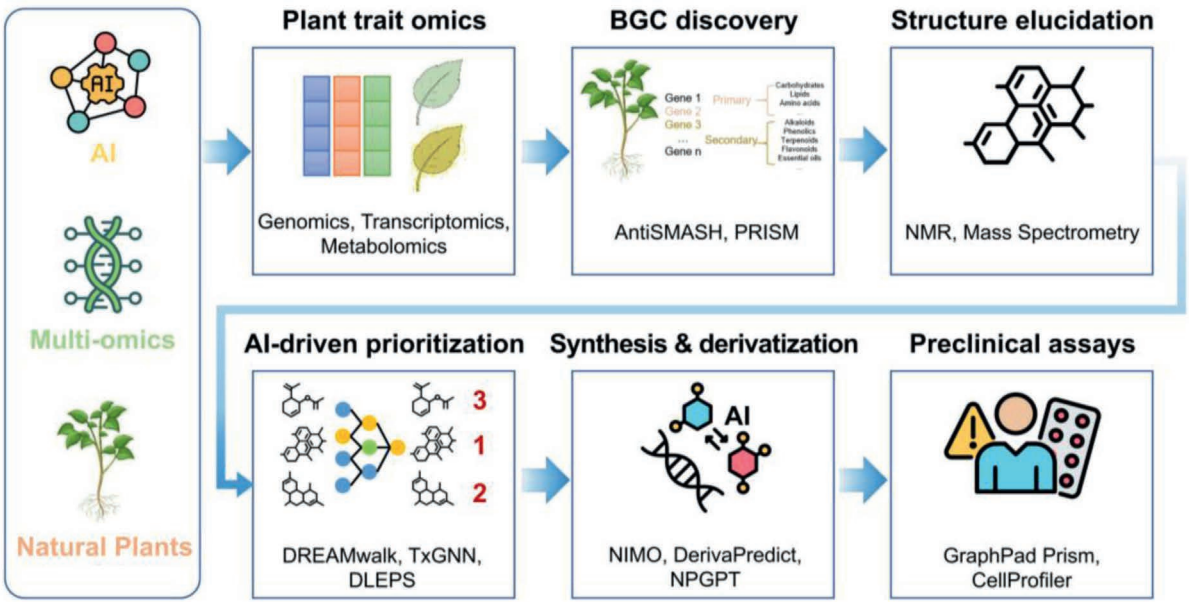

Figure 5 AI-driven multi-omics pipeline for plant-derived drug discovery.

transcriptional, epigenetic, and metabolic regulation[190]. Furthermore, combining genomic selection with environment-dependent pan-omics and ML within frameworks such as Green Systems Biology offers a promising avenue to improve medicinal plant breeding for enhanced yield, stress tolerance, and bioactive compound production, supported by rapid functional validation techniques like genome editing and speed breeding[191]. Together, these advances usher in a new era for natural product research and medicinal plant development.

The *Panax bipinnatifidus* complex, widespread in the Sino-Himalaya region, was clarified into four genetically distinct clades using ddRAD-seq data, revealing localized admixture and introgression, which complicate species identification and conservation[192]. In *Panax ginseng*, red skin root syndrome was linked to changes in primary and secondary metabolism, with increased anthocyanins and carotenoids, suggesting the role of jasmonate signaling in regulating this syndrome[193]. A high-density genetic map of Gardenia identified key QTLs for growth and leaf traits, aiding in the selection of desirable traits for future breeding[194]. In *Cryptomeria fortunei*, a multi-omics approach revealed that miRNA-target regulation controls terpenoid biosynthesis, with evergreen mutants maintaining high terpenoid levels during winter[195]. Liriope spicata's freezing tolerance was found to involve a complex regulatory network of flavonoids, ABA, and nitric oxide, crucial for osmotic adjustment and stress responses[196]. Finally, in Catharanthus roseus, a multi-omics approach identified key genes and cell-specific pathways in the biosynthesis of monoterpene indole alkaloids, important for cancer treatment[197].

3.4.2. AI In identification of medicinal plants

Medicinal plants play a crucial role in global health and pharmaceutical industries due to their therapeutic benefits and lower side effects compared to synthetic drugs. However, reliable identification and classification of these plants remain challenging, traditionally relying on expert botanists and time-consuming processes. Recent advances in AI, particularly deep learning techniques such as CNNs, have significantly improved automated medicinal plant recognition, achieving over 90% accuracy on key plant organs like leaves and flowers. Despite successes, challenges persist including performance variability due to species diversity, plant states, and limited availability of large, diverse, publicly accessible datasets. Transfer learning, data augmentation, and ensemble models have emerged as effective strategies to mitigate these issues, while mobile applications offer promising deployment avenues albeit lacking robust continuous refinement mechanisms[198,199]. Additionally, ML models have shown potential for early plant disease detection, further supporting plant health and productivity. Addressing current research gaps requires collaborative efforts to develop context-aware, efficient, and generalizable AI systems, ultimately enhancing biodiversity preservation and accelerating pharmaceutical research[200].

Recent advances in deep learning architectures have significantly improved medicinal plant classification. Hybrid models combining spectral-spatial 3D-CNNs with 2D-CNNs effectively capture complex hyperspectral image features, balancing computational efficiency and accuracy[201]. Vision Transformers leverage self-attention to model global image relationships, outperforming traditional CNNs in accuracy and robustness under diverse environmental conditions, enabling real-time applications on mobile devices[202]. In practical classification tasks, CNNs coupled with transfer learning techniques like MobileNetV2 and VGG16, demonstrate high accuracy on large medicinal leaf datasets, surpassing traditional ML methods like SVM and KNN which often lack scalability and struggle with complex leaf textures[203-205].

Besides, early disease detection is essential for plants' health and yield, but deep learning models require high-quality annotated datasets, which are costly and scarce. Embracing limited and imperfect datasets by categorizing and addressing their challenges facilitates practical deployment of plant disease recognition systems[206]. Image segmentation methods based on indices histograms have achieved over 92% accuracy in detecting diseased leaf regions in complex backgrounds[207]. Distinguishing medicinal from poisonous plants is critical due to public health risks. Deep learning models enhanced with spatial and channel attention mechanisms and diverse pooling strategies deliver over 99% accuracy, providing fast, non-destructive, and reliable identification alternatives to costly, error-prone traditional lab methods[208].

3.4.3. AI in synthetic biology of medicinal plants

Synthetic biology, with its vast applications across health, industry, agriculture, and the environment, is emerging as a key driver for sustainable solutions to modern societal challenges. The integration of AI is accelerating this field, enabling rapid advancements in plant-based technologies, such as protein engineering, phytochemical production, and microbiome engineering. These innovations are already leading to disruptive applications, including the enzymatic synthesis of new molecules, bioelectricity generation, and biomass-based construction materials. As the field continues to grow, synthetic biologists are poised to help achieve the vision of a sustainable, bio-based society in the coming decades[209]. A critical component of this progress is the elucidation of plant biosynthetic pathways, which enables the sustainable production of complex natural products. While genomic, transcriptomic, and metabolomic approaches have advanced significantly, much of the multi-omics data remains underutilized. To address this, AI-driven strategies are being developed to integrate molecular networking, reaction pair analysis, and gene expression patterns, enhancing the utilization of these data. Such advancements are expected to revolutionize pathway discovery, accelerating our understanding of plant metabolism and enabling the efficient bioengineering of valuable natural products[210].

Synthetic biology enables the redesign of biological systems for the synthesis of novel compounds, expanding chemical diversity beyond natural limitations. For instance, engineering yeast to synthesize noncanonical C11 terpenoids has enabled the production of new terpene scaffolds, expanding the chemical space of terpenoids[211]. Similarly, engineering orthogonal methyltransferases to catalyze carboxymethylation of substrates offers a new pathway for bioalkylation, increasing catalytic activity and selectivity[212]. The production of valuable chemicals is often constrained by native metabolic pathways, as seen with monoterpene biosynthesis in *Saccharomyces cerevisiae*, where a synthetic orthogonal pathway was established using neryl diphosphate to improve monoterpene production efficiency[213]. Meanwhile, directed evolution of glycosyltransferases has enabled regioselective glycosylation of flavonoids, enhancing the pharmaceutical potential of these compounds[214]. In the context of improving thiamine synthesis, the directed evolution of sulfide-dependent thiazole synthases in yeast demonstrated the potential to adapt anaerobic enzymes for efficient aerobic operation, further advancing synthetic biology applications in plant systems[215]. Additionally, the complete biosynthesis of cannabinoids in *Saccharomyces cerevisiae* from simple sugars like galactose opens

new possibilities for cannabinoid production and analog creation for medicinal purposes[216]. Furthermore, microbial biosynthesis of vinblastine, a potent anticancer drug, was achieved by refactoring the complex pathway in yeast, highlighting the potential of engineered microbes to produce natural and novel metabolites for industrial applications[217]. Also, elucidation of the biosynthesis of plant saponin adjuvants in the soapbark tree presents opportunities to optimize the production of immunostimulatory compounds for vaccine development[218]. Lastly, the discovery of the previously unknown enzyme mechanism in *Taxus mairei* for paclitaxel biosynthesis paves the way for the heterologous production of baccatin III, a key intermediate for the anticancer drug[219]. Similarly, expanding the chemical space of isoprenoids by synthesizing non-canonical 16-carbon terpenoids provides access to new compounds with potential applications in flavors and fragrances[220].

3.4.4. In traditional Chinese medicine

TCM offers unique advantages in disease prevention and treatment, but its holistic approach presents challenges when applying modern medical research models focused on single drugs and targets[221]. Network pharmacology provides a more suitable framework, linking drugs and diseases through biological networks to overcome the reductionist model, particularly in TCM research[222-224]. Recent advancements in network pharmacology, combined with AI and multi-omics technologies, have enriched this methodology, enabling a more systematic approach to understanding TCM's complex mechanisms[225-227]. AI-driven network target identification, coupled with multi-omics data, has proven valuable in analyzing TCM formulas, syndromes, and toxicity, enhancing the study of its bioactive components and treatment mechanisms. These innovations have accelerated the integration of TCM into modern medicine and promoted its scientific understanding and clinical applications. However, TCM-NP still faces challenges related to data quality and the need for more accurate algorithms and larger datasets. Future breakthroughs will rely on the continued development of AI and network pharmacology techniques, in which UNIQ system[228-230] offers as a representative system consisting of AI, multi-omics and network pharmacology, offering new opportunities for TCM's modernization and precision treatment.

Target identification is critical in drug research, especially for herbs, as it uncovers action mechanisms and reveals new therapeutic targets. The deep learning-based framework HTINet2, which combines TCM and clinical knowledge graphs with a residual graph convolution network, significantly enhances target prediction, as demonstrated in case studies with *Artemisia annua* and *Coptis chinensis*[231]. The complexity of herbal medicine makes it challenging to analyze global effects on the human body. The TMNP platform, using transcriptome-based network pharmacology, explores multi-scale effects of herbs, as exemplified by *Astragalus membranaceu*s and Xuesaitong injection, providing a more holistic approach than traditional chemical component-focused models[232]. Network pharmacology has gained traction in TCM by offering a systems-based understanding of herb−target-disease relationships. The integration of AI with Network pharmacology has advanced its application, improving predictions of herb-symptom treatments, as shown in TCM's treatment principles and therapeutic predictions for specific symptoms[229]. In precision medicine, single-drug treatments often fall short in addressing complex diseases. The SETComp model, leveraging transfer learning, improves predictions of complex system-cell-gene associations, demonstrating its potential in repositioning and compound synergy discovery[233].

The lack of analytical methods has historically hindered the application of TCM, but modern AI systems represented by UNIQ system, together with biology-driven omics approaches now offer new insights. Precision-modification metabolomics and chinmedomics are promising strategies for unveiling the complexity of TCM, enhancing its ability to treat diverse diseases effectively[234]. In chronic gastritis, the classification of Cold and Hot ZHENG syndromes remains underexplored at the molecular level. Transcriptomic analysis identified biomarkers for each syndrome, revealing Hot ZHENG as immune-driven and Cold ZHENG as immune-suppressive, with findings confirmed by validation using Jin Hong tablets[235]. Weifuchun Capsule (WFC) demonstrated its immunoregulatory effect on chronic atrophic gastritis by targeting the Toll-like receptor pathway, mediating HES6 expression. This was validated through *in vitro* and rat model experiments, confirming WFC's role in regulating immune responses and reducing inflammation[236].

4. Natural product databases enabling AI-driven discovery

Natural product databases play a crucial role in advancing AI-driven research by providing the essential data for training, validation, and integration in various multi-omics workflows. For instance, researchers have utilized the COCONUT database[237], combining machine learning techniques with molecular descriptors to screen for anti-aging drugs[238]. Another study introduced the DFT_ANPD method[239], which leverages molecular structure information from SMILES strings and semantic embeddings generated by large language models on datasets like NPACT[240] and NPASS[241] to predict the anticancer properties of natural compounds. Additionally, the KDHR[242] method utilized databases such as TCMSP[243] and ETCM[244] to construct a traditional Chinese medicine knowledge graph, integrating herb attribute information with multi-layer features from graph convolutional networks to propose a knowledge-driven herb recommendation system. These examples highlight how such databases can effectively support AI-driven drug discovery and natural product research.

4.1. Databases for natural compounds

Natural compounds databases provide critical chemical and biological information that supports drug discovery and biological research. These databases encompass detailed structures, bioactivity data, species sources, and therapeutic targets, significantly aiding in the screening and validation processes within pharmacological research (as shown in Table 4[237, 240-241, 245-250]). In addition to plant-derived compounds, these databases also cover natural products derived from microbes and marine organisms.

4.2. Databases for natural plants

Natural plant databases integrate diverse omics data, including genomic, transcriptomic, and metabolic information, facilitating comparative and functional genomics research. By providing comprehensive insights into plant biology and bioactive pathways, these resources (as shown in Table 5[251-255]) enhance the development of plant-derived medicines and deepen understanding of plant-based therapeutics.

4.3. Databases for traditional Chinese medicine

TCM databases systematically compile knowledge about medicinal herbs, herbal formulations, their constituent ingredients, bioactivities, clinical evidence, and therapeutic targets. These resources bridge traditional empirical knowledge with modern scientific validation, highlighting medicinal herbs as a significant component of medicinal plants and supporting their exploration in contemporary biomedical research and therapeutic innovation (as shown in Table 6[243-244, 256-258]).

5. Discussion and conclusions

5.1. Conclusions

Artificial intelligence and multi-omics approaches are significantly enhancing drug discovery by optimizing the identification of lead compounds and accelerating the entire development process. AI technologies, including ML and deep learning, enable more efficient data analysis, predictive modeling, and *de novo* drug design. These innovations have shifted drug discovery from traditional trial-and-error methods to data-driven, holistic approaches, improving precision and reducing timelines and costs[259,260]. By leveraging deep generative models and other AI tools, researchers can create novel chemical entities and optimize their pharmacological properties, moving faster through the drug development pipeline[261].

In the realm of natural product drug discovery, AI's role is particularly transformative. Natural products, which have long been a source of biologically active compounds, can now be explored more effectively through AI-driven multi-omics technologies. Advances in metabolomics, genomics, and molecular network analysis have enabled more efficient profiling and prioritization of natural extracts, revealing valuable insights into their bioactive compounds. The integration of AI in this field not only accelerates the discovery of potential therapeutic agents but also enhances the identification of key natural products by linking taxonomic and bioactivity data, thus opening new avenues for drug development[262,263]. This synergy between AI and multi-omics approaches is poised to greatly expedite the exploration of natural products, advancing the discovery of new, effective therapies.

5.2. Limitations in present studies

Currently, AI algorithms face certain limitations in both natural compounds and natural plants, especially in the development of medicinally valuable natural products. Research on natural compounds is currently a hot topic in the study of natural products. However, there are still some potential issues. It is well known that the performance of AI models is determined by a combination of factors such as the model structure, model parameters, and the quality and quantity of the data. In the case of natural compounds, some well-known natural compound-related literature and corresponding data, specifically the input-output pairs in the models, are excessively abundant, leading to a potential imbalance in the data during training. This may lead to the model's predictions being biased towards famous natural compounds with abundant related training data, weakening its ability to discover potentially valuable compounds or leading to incorrect predictions for these compounds. Data balancing methods may need to be applied to address such issues, in order to enhance the model's generalization and robustness.

The limitations of AI research related to natural plants mainly lie in the relative lack of various types of data. For example, multi-omics data of natural plants currently only involve a few hundred plant species, making it difficult to encompass the diverse multi-level omics characteristics of various natural plants. This further limits the development of models based on deep learning or Large Language Models (LLMs). With the accumulation of multi-omics technologies, larger-scale models will become feasible for algorithm development related to natural plants, which is also a key issue for future research.

Apart from the data scarcity and imbalance in model construction, there are also several other critical factors that may influence the robustness and accuracy of natural products data in drug development. For example, data leakage and assay-condition bias are significant challenges in natural products datasets, as experimental conditions and data handling can introduce unintended correlations that may mislead model predictions. Besides, the domain shift between natural products space and drug-like chemical space poses a challenge in transferring models trained on natural products to conventional drug-like molecules. Natural products often exhibit unique structural features that do not fully align with traditional drug-design criteria, making the prediction of their bioactivity more complex. In addition, the uncertainty in stereochemistry and 3D conformers within training data and generative models must be addressed, as incorrect stereochemical representations can significantly impact model performance, especially for compounds with chiral centers. Then, there are curation and coverage gaps in natural products and TCM databases, which may result in incomplete or biased data, limiting the generalizability and applicability of AI models. Finally, licensing, intellectual property, and indigenous knowledge ethics are essential considerations when mining traditional sources, as the utilization of indigenous knowledge must respect local laws, ensure proper attribution, and avoid exploitation of traditional communities. Addressing these issues is critical for advancing the ethical and effective application of AI in natural product-based drug discovery.

5.3. Future trends and prospects

Due to the current limitations, the development and application of AI algorithms in natural products is both highly challenging and promising. With the arrival of the LLMs era, an increasing number of the applications of LLMs and large deep learning models are emerging in the research of natural compounds and natural plants. Massive amounts of research literature are the types of samples that these large deep learning models excel at learning from. This enables the identification of natural products with medicinal value from natural compounds and plants, or the repurposing of existing medicinal natural products. In the near future, in the research of natural compounds combined with AI, due to the accumulation of a certain amount of data, there may be a shift from traditional structure-based models to LLMs-based models, as well as the development of AI agents specifically designed for natural compounds.

As for natural plants, since the field is relatively farther from drug development compared to natural compounds and lacks sufficient data, the trend should be twofold. On one hand, there will be a large output, accumulation, and updating of multi-omics data on natural plants, along with the establishment of related

Table 4 Databases for natural compounds.

Name	Description	URL	Statistics	Accessibility	Version/ Last updated	Fit for purpose	Ref.
NPACT	A curated database of plant derived natural compounds that exhibit anti-cancerous activity.	http://crdd.osdd.net/raghava/npact/	1574 bio-active compounds having activity against 353 cancer cell lines which corresponds to 5214 in-vitro compound-cell line interactions.	Open-access	2012-11	Anticancer natural compounds	240
NP-MRD	A database containing NMR spectra and structure data for natural compounds.	https://np-mrd.org/	281,859 compounds and 19,045 experimental NMR spectra.	Open-access	Version 2.0	NMR spectra	245
GNDC	A database dedicated to cataloging diverse natural components.	https://cbcb.cdutcm.edu.cn/gndc/	2.32 million secondary metabolites, 229 million small peptides, 2.38 million small RNAs, 0.26 million carbohydrates.	Open-access	Version 1.0	Natural components catalog	246
NPASS	A specialized database integrating natural products with their species sources and linking them to biological targets through experimental activity data.	https://bidd.group/NPASS/	94,413 unique natural products isolated from 32,287 source organisms and together with 958,866 activity records on 7753 targets.	Open-access	Version 2.0	Natural compounds' activity and species sources	241
CMAUP	A database of collective molecular activities of useful plants.	https://bidd.group/CMAUP/	11,472 experimental-derived activity values of 3029 plant ingredients against ~758 protein targets, 41,205 plant-ingredient pairs of ~60,222 ingredients of 7865 plants.	Open-access	Version 2.0	Collective molecular activities	247
COCONUT	A comprehensive platform facilitating natural product research by providing data, tools, and services for deposition, curation, and reuse.	https://coconut.naturalproducts.net/	695,117 molecules, 64 collections, 53,092 organisms, 35,185 citations.	Open-access	2025-09	Comprehensive natural compounds dataset	237
The natural products atlas	A database of microbially derived natural products.	https://www.npatlas.org/	1347 papers, 36,545 compounds.	Open-access	2024-09	Microbially-derived natural compounds	248
FOODB	A database of resource on food constituents, chemistry, and biology.	https://foodb.ca/	797 foods, 70,926 compounds, 5,150,056 contents, 162 pathways.	Open-access	Version 1.0	Food compound database	249
CMNPD	A comprehensive marine natural products database based on manually curated data.	https://www.cmnpd.org/	31,561 compounds, 3354 organisms, 2652 targets, 72,349 bioactivities, 128,488 documents.	Open-access	2020-09	Marine natural compounds database	250

databases. Data serves as the foundation for AI to develop algorithms and applications in natural plant research. On the other hand, with the accumulation of research and literature on natural plants, including studies on compound synthesis and metabolism in natural plants, identification of natural compounds, and multi-omics research on natural compounds, a certain data foundation will be provided. This will lead to the application of LLMs-based models in natural plant research, offering knowledge inference from a natural language perspective.

Whether it is natural plants or natural compounds, drug development based on natural products is an important component of the biomedical field. In the future, the trend of AI-based drug development for natural products will lean towards the generation and accumulation of more data, the establishment of higher-quality databases, and the creation of models with greater accuracy. Due to its unique advantages, LLMs models, which can learn and infer from large volumes of relevant research, may play a key role in the future of AI-driven drug development for natural products.

Table 5 Databases for natural plants.

Name	Description	URL	Statistics	Accessibility	Version/ Last updated	Fit for purpose	Ref.
WP-MOD	A database integrating taxonomic and multi-omics data for 373 woody plant species.	https://www.woodyplant.com/	Multi-omics information of 373 woody plants with whole-genome references, transcriptomes, methylome, small RNA transcriptomes, degradome, epigenome, etc.	Open-access	Version 1.1	Woody plant multi-omics	251
PMN	A plant metabolic pathway database.	https://plantcyc.org/	One multi-species reference database called PlantCyc and 155 species/taxon-specific databases.	Open-access	Version 16.0	Plant metabolome	252
GreenPhylDB	A web resource designed for comparative and functional genomics in plants.	https://www.greenphyl.org	19 pangenomes and 27 reference genomes for 46 species.	Open-access	Version 5.1	Plant genome	253
1 K-MPGD	A comprehensive set of omics data and KEGG pathway information for medicinal plants.	http://www.herbgenome.com/	113 genomes of medicinal plants.	Access granted upon request	2021-06	Medicinal plant genome	254
PlantscRNAdb	A plant single-cell transcriptome database.	http://ibi.zju.edu.cn/plantscrnadb/	231,939 candidate markers, 58,846 high confidence markers, 38,232 pathway markers and 5243 high confidence pathway markers from 33 species.	Open-access	Version 4.0	Plant scRNA	255

Table 6 Databases for TCM.

Name	Description	URL	Statistics	Accessibility	Version/ Last Updated	Fit for purpose	Ref.
TCMSP	A unique systems pharmacology platform of Chinese herbal medicines that captures the relationships between drugs, targets and diseases.	https://www.91tcmsp.com/	499 herbs registered in Chinese pharmacopoeia (2010), with a total of 12,144 chemicals.	Subscription-based	Version 2.3	TCM comprehensive database	243
TCMPG	An integrative database for storing the scattered genomes of medicinal plants.	http://cbcb.cdutcm.edu.cn/TCMPG/	160 medicinal plants, 195 corresponding genomes, and 255 herbal medicines.	Open-access	Version 2.0	TCM plant genome	256
HERB	A database integrating clinical and experimental evidences for TCM.	http://47.92.70.12/	6743 formulae, 6892 herbs, 44,595 ingredients, 15,515 targets, 30,170 diseases, 8558 clinical trials, 8032 meta-analyses, 6644 references, and 2231 expression profiles.	Open-access	Version 2.0	TCM comprehensive database	257
ETCM	An encyclopaedia of TCM.	http://www.tcmip.cn/ETCM2/front/#/	48,442 TCM formulas, 9872 Chinese patent drugs, 2079 Chinese medicinal materials and 38,298 ingredients.	Open-access	Version 2.0	TCM formula	244
TCMBank	A database with TCM, ingredients, diseases, and gene targets, and a comprehensive network containing six pairs of relationships.	https://tcmbank.cn/	9191 herbs, 61,965 components, 15,179 targets, and 32,529 diseases	Open-access	2023-03	TCM comprehensive network	258

5.4. Summary

In this review, we provide a comprehensive overview of the methods and applications of AI and multi-omics technologies in drug development, including natural compounds and natural plants. We also compile various databases that have the potential to provide data for the application of AI and multi-omics technologies in natural products.

Acknowledgments

This work was supported by the Innovation Team and Talent Support Program Project of Traditional Chinese Medicine, National Administration of Traditional Chinese Medicine (ZYYCXTD-D-202405, China), the National Natural Science Foundation of China (T2341008), the Pilot Project for Disciplinary Breakthroughs of Ministry of Education (Prevention and Treatment of Multi-System Comorbid Diseases with Traditional Chinese Medicine) and the Anhui Province Traditional Chinese Medicine Science and Technology Research Project (202303a07020001, China).

Author contributions

Boyang Wang: Writing − review & editing, Writing − original draft, Visualization. Qingyuan Liu: Writing − review & editing, Writing − original draft, Visualization. Tingyu Zhang: Writing − original draft. Weibo Zhao: Writing − original draft. Dingfan Zhang: Visualization. Chayanis Sutcharitchan: Writing − original draft. Shao Li: Writing − review & editing, Supervision, Conceptualization. All the authors have read and approved the final manuscript.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Atanasov AG, Zotchev SB, Dirsch VM, Supuran CT. Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov* 2021;**20**:200−16.
2. Lu X, Friedrich LJ, Efferth T. Natural products targeting tumour angiogenesis. *Br J Pharmacol* 2025;**182**:2094−136.
3. Wang L, Yan Y, Wu L, Peng J. Natural products in non-alcoholic fatty liver disease (NAFLD): novel lead discovery for drug development. *Pharmacol Res* 2023;**196**:106925.
4. Chopra B, Dhingra AK. Natural products:a lead for drug discovery and development. *Phytother Res* 2021;**35**:4660−702.
5. Wang Y, Wang F, Liu W, Geng Y, Shi Y, Tian Y, et al. New drug discovery and development from natural products: advances and strategies. *Pharmacol Ther* 2024;**264**:108752.
6. Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, et al. Scientific discovery in the age of artificial intelligence. *Nature* 2023;**620**: 47−60.
7. Zhang K, Yang X, Wang Y, Yu Y, Huang N, Li G, et al. Artificial intelligence in drug development. *Nat Med* 2025;**31**:45−59.
8. Elemento O, Leslie C, Lundin J, Tourassi G. Artificial intelligence in cancer research, diagnosis and therapy. *Nat Rev Cancer* 2021;**21**: 747−52.
9. Song AH, Jaume G, Williamson DFK, Lu MY, Vaidya A, Miller TR, et al. Artificial intelligence for digital and computational pathology. *Nat Rev Bioeng* 2023;**1**:930−49.
10. van Breugel B, Liu T, Oglic D, van der Schaar M. Synthetic data in biomedicine *via* generative artificial intelligence. *Nat Rev Bioeng* 2024;**2**:991−1004.
11. Lewin HA, Robinson GE, Kress WJ, Baker WJ, Coddington J, Crandall KA, et al. Earth biogenome project: sequencing life for the future of life. *Proc Natl Acad Sci U S A* 2018;**115**:4325−33.
12. Wang Y, Xiao Z, Wang Z, Lee D, Ma Y, Wilhelm S, et al. Multi-omics approaches to decipher the interactions of nanoparticles and biological systems. *Nat Rev Bioeng* 2025;**3**:333−48.
13. Haniffa M, Maartens A, Winheim E, Jardine L. Decoding the human prenatal immune system with single-cell multi-omics. *Nat Rev Immunol* 2025;**25**:285−97.
14. Baysoy A, Bai Z, Satija R, Fan R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* 2023; **24**:695−713.
15. Vandereyken K, Sifrim A, Thienpont B, Voet T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* 2023;**24**:494−515.
16. Gao F, Huang K, Xing Y. Artificial intelligence in omics. *Genom Proteom Bioinform* 2022;**20**:811−3.
17. Cui H, Tejada-Lapuerta A, Brbić M, Saez-Rodriguez J, Cristea S, Goodarzi H, et al. Towards multimodal foundation models in molecular cell biology. *Nature* 2025;**640**:623−33.
18. Steyaert S, Pizurica M, Nagaraj D, Khandelwal P, Hernandez-Boussard T, Gentles AJ, et al. Multimodal data fusion for cancer biomarker discovery with deep learning. *Nat Mach Intell* 2023;**5**: 351−62.
19. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med* 2022;**28**:1773−84.
20. Catacutan DB, Alexander J, Arnold A, Stokes JM. Machine learning in preclinical drug discovery. *Nat Chem Biol* 2024;**20**:960−73.
21. Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. *Nat Mach Intell* 2020;**2**:573−84.
22. Schork NJ, Goetz LH. From precision interventions to precision health. *Nat Commun* 2025;**16**:5024.
23. Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJJ, Martin NI, et al. Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov* 2023;**22**:895−916.
24. Cui S, Li Q, Li D, Lian Z, Hou J. Hyper-mol: molecular representation learning *via* fingerprint-based hypergraph. *Comput Intell Neurosci* 2023;**2023**:3756102.
25. Ilnicka A, Schneider G. Compression of molecular fingerprints with autoencoder networks. *Mol Inform* 2023;**42**:e2300059.
26. Hajiabolhassan H, Taheri Z, Hojatnia A, Yeganeh YT. Funqg: molecular representation learning *via* quotient graphs. *J Chem Inf Model* 2023;**63**:3275−87.
27. Li Y, Liu B, Deng J, Guo Y, Du H. Image-based molecular representation learning for drug development: a survey. *Briefings Bioinf* 2024;**25**:bbae294.
28. Xiang H, Jin S, Liu X, Zeng X, Zeng L. Chemical structure-aware molecular image representation learning. *Briefings Bioinf* 2023;**24**: bbad404.
29. Li C, Wei W, Li J, Yao J, Zeng X, Lv Z. 3Dmol-net: learn 3D molecular representation using adaptive graph convolutional network based on rotation invariance. *IEEE J Biomed Health Inform* 2022;**26**: 5044−54.
30. Lim J, Ryu S, Park K, Choe YJ, Ham J, Kim WY. Predicting drug−target interaction using a novel graph neural network with 3D structure-embedded graph representation. *J Chem Inf Model* 2019; **59**:3981−8.
31. Gui S, Chen Z, Lu B, Chen M. Molecular sparse representation by a 3D ellipsoid radial basis function neural network *via* l1 regularization. *J Chem Inf Model* 2020;**60**:6054−64.
32. Zhang Y, Bai X. Geometry-augmented molecular representation learning for property prediction. *IEEE ACM Trans Comput Biol Bioinf* 2024;**21**:1518−28.
33. Tang Q, Nie F, Zhao Q, Chen W. A merged molecular representation deep learning method for blood−brain barrier permeability prediction. *Briefings Bioinf* 2022;**23**:bbac357.
34. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, et al. Improved protein structure prediction using potentials from deep learning. *Nature* 2020;**577**:706−10.
35. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with alphafold. *Nature* 2021;**596**:583−9.
36. Zhao B, Xu W, Guan J, Zhou S. Molecular property prediction based on graph structure learning. *Bioinformatics* 2024;**40**:btae304.

37. Duan Y, Yang X, Zeng X, Wang W, Deng Y, Cao D. Enhancing molecular property prediction through task-oriented transfer learning: integrating universal structural insights and domain-specific knowledge. *J Med Chem* 2024;**67**:9575−86.
38. Yang X, Duan Y, Cheng Z, Li K, Liu Y, Zeng X, et al. Mpcd: a multitask graph transformer for molecular property prediction by integrating common and domain knowledge. *J Med Chem* 2024;**67**: 21303−16.
39. Schuh MG, Boldini D, Sieber SA. Synergizing chemical structures and bioassay descriptions for enhanced molecular property prediction in drug discovery. *J Chem Inf Model* 2024;**64**:4640−50.
40. Jiang T, Wang Z, Yu W, Wang J, Yu S, Bao X, et al. Mix-key: graph mixup with key structures for molecular property prediction. *Briefings Bioinf* 2024;**25**:bbae165.
41. Dalkıran A, Atakan A, Rifaioğlu AS, Martin MJ, Atalay R, Acar AC, et al. Transfer learning for drug−target interaction prediction. *Bioinformatics* 2023;**39**:i103−10.
42. Chen H, King FJ, Zhou B, Wang Y, Canedy CJ, Hayashi J, et al. Drug target prediction through deep learning functional representation of gene signatures. *Nat Commun* 2024;**15**:1853.
43. Atas Guvenilir H, Doğan T. How to approach machine learning-based prediction of drug/compound−target interactions. *J Cheminf* 2023;**15**:16.
44. Hu J, Yu W, Pang C, Jin J, Pham NT, Manavalan B, et al. Drugormerdti: drug graphormer for drug−target interaction prediction. *Comput Biol Med* 2023;**161**:106946.
45. Vatansever S, Schlessinger A, Wacker D, Kaniskan H, Jin J, Zhou MM, et al. Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: state-of-the-arts and future directions. *Med Res Rev* 2021;**41**:1427−73.
46. Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. *Cell* 2020;**180**:688−702.e13.
47. Noor F, Junaid M, Almalki AH, Almaghrabi M, Ghazanfar S, Tahir Ul Qamar M. Deep learning pipeline for accelerating virtual screening in drug discovery. *Sci Rep* 2024;**14**:28321.
48. Liu Q, Chen Z, Wang B, Pan B, Zhang Z, Shen M, et al. Leveraging network target theory for efficient prediction of drug−disease interactions: a transfer learning approach. *Adv Sci (Weinh)* 2025;**12**: e2409130.
49. Wang S, Song T, Zhang S, Jiang M, Wei Z, Li Z. Molecular substructure tree generative model for *de novo* drug design. *Briefings Bioinf* 2022;**23**:bbab592.
50. Tong X, Liu X, Tan X, Li X, Jiang J, Xiong Z, et al. Generative models for *de novo* drug design. *J Med Chem* 2021;**64**:14011−27.
51. Zhang Z, Liu Q, Lee CK, Hsieh CY, Chen E. An equivariant generative framework for molecular graph-structure co-design. *Chem Sci* 2023;**14**:8380−92.
52. Wang J, Chu Y, Mao J, Jeon HN, Jin H, Zeb A, et al. *De novo* molecular design with deep molecular generative models for PPI inhibitors. *Briefings Bioinf* 2022;**23**:bbac285.
53. Zhung W, Kim H, Kim WY. 3D molecular generative framework for interaction-guided drug design. *Nat Commun* 2024;**15**:2688.
54. Chen S, Noh J, Jang J, Kim S, Gu GH, Jung Y. Reaction templates: bridging synthesis knowledge and artificial intelligence. *Acc Chem Res* 2024;**57**:1964−72.
55. Long L, Li R, Zhang J. Artificial intelligence in retrosynthesis prediction and its applications in medicinal chemistry. *J Med Chem* 2025;**68**:2333−55.
56. Dong J, Zhao M, Liu Y, Su Y, Zeng X. Deep learning in retrosynthesis planning: datasets, models and tools. *Briefings Bioinf* 2022;**23**: bbab391.
57. Zhu J, Wang J, Wang X, Gao M, Guo B, Gao M, et al. Prediction of drug efficacy from transcriptional profiles with deep learning. *Nat Biotechnol* 2021;**39**:1444−52.
58. Wu Y, Chen M, Qin Y. Anticancer drug response prediction integrating multi-omics pathway-based difference features and multiple deep learning techniques. *PLoS Comput Biol* 2025;**21**:e1012905.
59. Wang Y, Yang Y, Chen S, Wang J. Deepdrk: a deep learning framework for drug repurposing through kernel-based multi-omics integration. *Briefings Bioinf* 2021;**22**:bbab048.
60. Choi J, Lee J. V-dock: fast generation of novel drug-like molecules using machine-learning-based docking score and molecular optimization. *Int J Mol Sci* 2021;**22**:11635.
61. Bande AY, Baday S. Accelerating molecular docking using machine learning methods. *Mol Inform* 2024;**43**:e202300167.
62. Yu L, He X, Fang X, Liu L, Liu J. Deep learning with geometry-enhanced molecular representation for augmentation of large-scale docking-based virtual screening. *J Chem Inf Model* 2023;**63**:6501−14.
63. Yin Y, Lam HYI, Mu Y, Li HY, Kong AW. Advancing bioactivity prediction through molecular docking and self-attention. *IEEE J Biomed Health Inform* 2024;**28**:7599−610.
64. Shim H, Kim H, Allen JE, Wulff H. Pose classification using three-dimensional atomic structure-based neural networks applied to ion channel-ligand docking. *J Chem Inf Model* 2022;**62**:2301−15.
65. Shi W, Singha M, Pu L, Srivastava G, Ramanujam J, Brylinski M. Graphsite: ligand binding site classification with deep graph learning. *Biomolecules* 2022;**12**:1053.
66. Ketata MA, Laue C, Mammadov R, Stärk H, Wu M, Corso G, et al. Diffdock-pp: rigid protein−protein docking with diffusion models. *Machine Learning for Drug Discovery workshop, ICLR* 2023.
67. Ji B, He X, Zhai J, Zhang Y, Man VH, Wang J. Machine learning on ligand−residue interaction profiles to significantly improve binding affinity prediction. *Briefings Bioinf* 2021;**22**:bbab054.
68. Wang S, Liu D, Ding M, Du Z, Zhong Y, Song T, et al. Se-onionnet: a convolution neural network for protein−ligand binding affinity prediction. *Front Genet* 2020;**11**:607824.
69. Rezaei MA, Li Y, Wu D, Li X, Li C. Deep learning in drug design: protein−ligand binding affinity prediction. *IEEE ACM Trans Comput Biol Bioinf* 2022;**19**:407−17.
70. Zhang X, Gao H, Wang H, Chen Z, Zhang Z, Chen X, et al. Planet: a multi-objective graph neural network model for protein−ligand binding affinity prediction. *J Chem Inf Model* 2024;**64**:2205−20.
71. Morehead A, Cheng J. Flowdock: geometric flow matching for generative protein−ligand docking and affinity prediction. *Bioinformatics* 2025;**41**:i198−206.
72. Liu F, Xu H, Cui P, Li S, Wang H, Wu Z. Nfsa-dti: a novel drug−target interaction prediction model using neural fingerprint and self-attention mechanism. *Int J Mol Sci* 2024;**25**:11818.
73. Li F, Zhang Z, Guan J, Zhou S. Effective drug−target interaction prediction with mutual interaction neural network. *Bioinformatics* 2022;**38**:3582−9.
74. Bian J, Zhang X, Zhang X, Xu D, Wang G. Mcanet: shared-weight-based multiheadcrossattention network for drug−target interaction prediction. *Briefings Bioinf* 2023;**24**:bbad082.
75. Liu S, Wang Y, Deng Y, He L, Shao B, Yin J, et al. Improved drug−target interaction prediction with intermolecular graph transformer. *Briefings Bioinf* 2022;**23**:bbac162.
76. Zhao T, Hu Y, Valsdottir LR, Zang T, Peng J. Identifying drug−target interactions based on graph convolutional network and deep neural network. *Briefings Bioinf* 2021;**22**:2141−50.
77. Tian Z, Yu Y, Ni F, Zou Q. Drug−target interaction prediction with collaborative contrastive learning and adaptive self-paced sampling strategy. *BMC Biol* 2024;**22**:216.
78. Li Y, Qiao G, Gao X, Wang G. Supervised graph co-contrastive learning for drug−target interaction prediction. *Bioinformatics* 2022;**38**:2847−54.
79. Wei J, Zhu Y, Zhuo L, Liu Y, Fu X, Li F. Efficient deep model ensemble framework for drug−target interaction prediction. *J Phys Chem Lett* 2024;**15**:7681−93.
80. Huang K, Chandak P, Wang Q, Havaldar S, Vaid A, Leskovec J, et al. A foundation model for clinician-centered drug repurposing. *Nat Med* 2024;**30**:3601−13.
81. Bang D, Lim S, Lee S, Kim S. Biomedical knowledge graph learning for drug repurposing by extending guilt-by-association to multiple layers. *Nat Commun* 2023;**14**:3570.

82. Ohnuki Y, Akiyama M, Sakakibara Y. Deep learning of multimodal networks with topological regularization for drug repositioning. *J Cheminf* 2024;**16**:103.
83. Zhang H, Cui H, Zhang T, Cao Y, Xuan P. Learning multi-scale heterogenous network topologies and various pairwise attributes for drug−disease association prediction. *Briefings Bioinf* 2022;**23**: bbac009.
84. Zhao BW, Su XR, Hu PW, Ma YP, Zhou X, Hu L. A geometric deep learning framework for drug repositioning over heterogeneous information networks. *Briefings Bioinf* 2022;**23**:bbac384.
85. Meng Y, Lu C, Jin M, Xu J, Zeng X, Yang J. A weighted bilinear neural collaborative filtering approach for drug repositioning. *Briefings Bioinf* 2022;**23**:bbab581.
86. Li Z, Huang Q, Chen X, Wang Y, Li J, Xie Y, et al. Identification of drug−disease associations using information of molecular structures and clinical symptoms *via* deep convolutional neural network. *Front Chem* 2019;**7**:924.
87. Amiri R, Razmara J, Parvizpour S, Izadkhah H. A novel efficient drug repurposing framework through drug−disease association data integration using convolutional neural networks. *BMC Bioinf* 2023; **24**:442.
88. Gao Y, Zhang X, Sun Z, Chandak P, Bu J, Wang H. Precision adverse drug reactions prediction with heterogeneous graph neural network. *Adv Sci (Weinh)* 2024;**12**:e2404671.
89. Xuan P, Wang M, Liu Y, Wang D, Zhang T, Nakaguchi T. Integrating specific and common topologies of heterogeneous graphs and pairwise attributes for drug-related side effect prediction. *Briefings Bioinf* 2022;**23**:bbac126.
90. Xuan P, Cheng T, Cui H, Gu J, Jin Q, Zhang T. Multi-knowledge graph and multi-view entity feature learning for predicting drug-related side effects. *J Chem Inf Model* 2025;**65**: 5124−38.
91. Xuan P, Li P, Cui H, Wang M, Nakaguchi T, Zhang T. Learning multi-types of neighbor node attributes and semantics by heterogeneous graph transformer and multi-view attention for drug-related side-effect prediction. *Molecules* 2023;**28**:6544.
92. Yu Z, Wu Z, Li W, Liu G, Tang Y. Adenet: a novel network-based inference method for prediction of drug adverse events. *Briefings Bioinf* 2022;**23**:bbab580.
93. Lin J, He Y, Ru C, Long W, Li M, Wen Z. Advancing adverse drug reaction prediction with deep chemical language model for drug safety evaluation. *Int J Mol Sci* 2024;**25**:4516.
94. Cheng X, Cheng M, Yu L, Xiao X. Iadrgse: a graph-embedding and self-attention encoding for identifying adverse drug reaction in the earlier phase of drug development. *Int J Mol Sci* 2022;**23**: 16216.
95. Xu X, Yue L, Li B, Liu Y, Wang Y, Zhang W, et al. Dsgat: predicting frequencies of drug side effects by graph attention networks. *Briefings Bioinf* 2022;**23**:bbab586.
96. Zhao H, Li D, Zhong J, Liang X, Duan G, Wang J. A deep learning-based method for predicting the frequency classes of drug side effects based on multi-source similarity fusion. *Bioinformatics* 2025;**41**: btaf319.
97. Park S, Lee S, Pak M, Kim S. Dual representation learning for predicting drug-side effect frequency using protein target information. *IEEE J Biomed Health Inform* 2025;**29**:1817−27.
98. Uner OC, Kuru HI, Cinbis RG, Tastan O, Cicek AE. Deepside: a deep learning approach for drug side effect prediction. *IEEE ACM Trans Comput Biol Bioinf* 2023;**20**:330−9.
99. Wang Z, Dong J, Wu L, Dai C, Wang J, Wen Y, et al. Deml: drug synergy and interaction prediction using ensemble-based multi-task learning. *Molecules* 2023;**28**:844.
100. Rafiei F, Zeraati H, Abbasi K, Ghasemi JB, Parsaeian M, Masoudi-Nejad A. Deeptrasynergy: drug combinations using multimodal deep learning with transformers. *Bioinformatics* 2023;**39**:btad438.
101. Yang J, Xu Z, Wu WKK, Chu Q, Zhang Q. Graphsynergy: a network-inspired deep learning model for anticancer drug combination prediction. *J Am Med Inf Assoc* 2021;**28**:2336−45.
102. Zhang G, Gao Z, Yan C, Wang J, Liang W, Luo J, et al. Kgansynergy: knowledge graph attention network for drug synergy prediction. *Briefings Bioinf* 2023;**24**:bbad167.
103. Jin Q, Zhang X, Huo D, Xie H, Zhang D, Liu L, et al. Predicting drug synergy using a network propagation inspired machine learning framework. *Brief Funct Genom* 2024;**23**:429−40.
104. Al-Rabeah MH, Lakizadeh A. Prediction of drug−drug interaction events using graph neural networks based feature extraction. *Sci Rep* 2022;**12**:15590.
105. Tang YC, Gottlieb A. Synpathy: predicting drug synergy through drug-associated pathways using deep learning. *Mol Cancer Res* 2022;**20**:762−9.
106. Zhang F, Zhao X, Wei J, Wu L. Pathsynergy: a deep learning model for predicting drug synergy in liver cancer. *Briefings Bioinf* 2025;**26**: bbaf192.
107. Kim E, Nam H. Deside-DDI: interpretable prediction of drug−drug interactions using drug-induced gene expressions. *J Cheminf* 2022; **14**:9.
108. Nyamabo AK, Yu H, Shi JY. SSI-DDI: substructure−substructure interactions for drug−drug interaction prediction. *Briefings Bioinf* 2021;**22**:bbab133.
109. He H, Chen G, Yu-Chian Chen C. 3DGT-DDI: 3D graph and text based neural network for drug−drug interaction prediction. *Briefings Bioinf* 2022;**23**:bbac134.
110. Atance SR, Diez JV, Engkvist O, Olsson S, Mercado R. *De novo* drug design using reinforcement learning with graph-based deep generative models. *J Chem Inf Model* 2022;**62**:4863−72.
111. Mukaidaisi M, Vu A, Grantham K, Tchagang A, Li Y. Multi-objective drug design based on graph-fragment molecular representation and deep evolutionary learning. *Front Pharmacol* 2022;**13**:920747.
112. Li Y, Pei J, Lai L. Structure-based *de novo* drug design using 3D deep generative models. *Chem Sci* 2021;**12**:13664−75.
113. Yang Y, Hsieh CY, Kang Y, Hou T, Liu H, Yao X. Deep generation model guided by the docking score for active molecular design. *J Chem Inf Model* 2023;**63**:2983−91.
114. Zhang J, Chen H. *De novo* molecule design using molecular generative models constrained by ligand−protein interactions. *J Chem Inf Model* 2022;**62**:3291−306.
115. Mao J, Wang J, Zeb A, Cho KH, Jin H, Kim J, et al. Transformer-based molecular generative model for antiviral drug design. *J Chem Inf Model* 2024;**64**:2733−45.
116. Nguyen T, Karolak A. Transformer graph variational autoencoder for generative molecular design. *Biophys J* 2025;**124**:3867−75.
117. Wang F, Feng X, Guo X, Xu L, Xie L, Chang S. Improving *de novo* molecule generation by embedding lstm and attention mechanism in cyclegan. *Front Genet* 2021;**12**:709500.
118. Zhu JY, Park T, Isola P, Efros AA. *Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017 IEEE international conference on computer vision (ICCV)*. 2017. p. 2242−51. Venice, Italy.
119. Jang H, Seo S, Park S, Kim BJ, Choi GW, Choi J, et al. *De novo* drug design through gradient-based regularized search in information-theoretically controlled latent space. *J Comput Aided Mol Des* 2024;**38**:32.
120. Runcie NT, Mey A. Silvr: guided diffusion for molecule generation. *J Chem Inf Model* 2023;**63**:5996−6005.
121. Oestreich M, Merdivan E, Lee M, Schultze JL, Piraud M, Becker M. Drugdiff: small molecule diffusion model with flexible guidance towards molecular properties. *J Cheminf* 2025;**17**:23.
122. Li C, Yao J, Wei W, Niu Z, Zeng X, Li J, et al. Geometry-based molecular generation with deep constrained variational autoencoder. *IEEE Transact Neural Networks Learn Syst* 2024;**35**:4852−61.
123. Song Y, Zhang M, Chang S, Chu G, Ji H. Derivapredict: a user-friendly tool for predicting and evaluating active derivatives of natural products. *Molecules* 2025;**30**:1683.
124. Shen X, Zeng T, Chen N, Li J, Wu R. Nimo: a natural product-inspired molecular generative model based on conditional transformer. *Molecules* 2024;**29**:1867.

125. Sakano K, Furui K, Ohue M. NPGPT: natural product-like compound generation with GPT-based chemical language models. *J Supercomput* 2024;**81**:352.
126. Tay DWP, Yeo NZX, Adaikkappan K, Lim YH, Ang SJ. 67 million natural product-like compound database generated *via* molecular language processing. *Sci Data* 2023;**10**:296.
127. Merk D, Grisoni F, Friedrich L, Schneider G. Tuning artificial intelligence on the *de novo* design of natural-product-inspired retinoid X receptor modulators. *Commun Chem* 2018;**1**:68.
128. Yang R, Zha X, Gao X, Wang K, Cheng B, Yan B. Multi-stage virtual screening of natural products against p38α; mitogen-activated protein kinase: predictive modeling by machine learning, docking study and molecular dynamics simulation. *Heliyon* 2022;**8**:e10495.
129. Liu Z, Huang D, Zheng S, Song Y, Liu B, Sun J, et al. Deep learning enables discovery of highly potent anti-osteoporosis natural products. *Eur J Med Chem* 2021;**210**:112982.
130. Friedrich L, Cingolani G, Ko YH, Iaselli M, Miciaccia M, Perrone MG, et al. Learning from nature: from a marine natural product to synthetic cyclooxygenase-1 inhibitors by automated *de novo* design. *Adv Sci (Weinh)* 2021;**8**:e2100832.
131. Wang P, Lehti-Shiu MD, Lotreck S, Segura Abá K, Krysan PJ, Shiu SH. Prediction of plant complex traits *via* integration of multi-omics data. *Nat Commun* 2024;**15**:6856.
132. Abdullah-Zawawi MR, Govender N, Harun S, Muhammad NAN, Zainal Z, Mohamed-Hussein ZA. Multi-omics approaches and resources for systems-level gene function prediction in the plant kingdom. *Plants* 2022;**11**:2614.
133. Sharma V, Gupta P, Priscilla K, SharanKumar, Hangargi B, Veershetty A, et al. Metabolomics intervention towards better understanding of plant traits. *Cells* 2021;**10**:346.
134. Zhan C, Shen S, Yang C, Liu Z, Fernie AR, Graham IA, et al. Plant metabolic gene clusters in the multi-omics era. *Trends Plant Sci* 2022;**27**:981−1001.
135. Srikanth P, Sivakumar D, Sharma A, Kaushik N. Recent developments in omics techniques for improving plant abiotic stress using microbes. *Int J Environ Sci Technol* 2025;**22**:3787−810.
136. Grossart HP, Massana R, McMahon KD, Walsh DA. Linking metagenomics to aquatic microbial ecology and biogeochemical cycles. *Limnol Oceanogr* 2020;**65**:S2−20.
137. Ijaz S, Iqbal J, Abbasi BA, Yaseen T, Rehman S, Kazi M, et al. Role of omics-based technologies in plant sciences. In: Fiaz S, Prakash CS, editors. *OMICs-based techniques for global food security.* John Wiley & Sons, Ltd.; 2024. p. 45−66.
138. Nasar S, Muhammad M, Siddiqi EH, Majeed M, Ameen N, Ullah S, et al. Omics-based knowledge for achieving food and nutritional security. In: Fiaz S, Prakash CS, editors. *OMICs-based techniques for global food security.* John Wiley & Sons, Ltd.; 2024. p. 67−90.
139. Rodríguez A, Castrejón-Godínez ML, Salazar-Bustamante E, Gama-Martínez Y, Sánchez-Salinas E, Mussali-Galante P, et al. Omics approaches to pesticide biodegradation. *Curr Microbiol* 2020;**77**:545−63.
140. Raza A. Metabolomics: a systems biology approach for enhancing heat stress tolerance in plants. *Plant Cell Rep* 2022;**41**:741−63.
141. Chandran H, Meena M, Sharma K. Microbial biodiversity and bioremediation assessment through omics approaches. *Front Environ Chem* 2020;**1**:2020.
142. Boycheva S, Daviet L, Wolfender JL, Fitzpatrick TB. The rise of operon-like gene clusters in plants. *Trends Plant Sci* 2014;**19**:447−59.
143. Okada A, Okada K, Miyamoto K, Koga J, Shibuya N, Nojiri H, et al. Ostgap1, a bzip transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. *J Biol Chem* 2009;**284**:26510−8.
144. Nützmann HW, Huang A, Osbourn A. Plant metabolic clusters−from genetics to genomics. *New Phytol* 2016;**211**:771−89.
145. Zhan C, Lei L, Liu Z, Zhou S, Yang C, Zhu X, et al. Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nat Plants* 2020;**6**:1447−54.
146. Shen S, Peng M, Fang H, Wang Z, Zhou S, Jing X, et al. An oryza-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. *Sci Bull (Beijing)* 2021;**66**:2369−80.
147. Jeon JE, Kim JG, Fischer CR, Mehta N, Dufour-Schroif C, Wemmer K, et al. A pathogen-responsive gene cluster for highly modified fatty acids in tomato. *Cell* 2020;**180**:176−87.e19.
148. Ji P, Yang X, Zhao X. Application of metabolomics in quality control of traditional chinese medicines: a review. *Front Plant Sci* 2024;**15**:1463666.
149. Liu FJ, Jiang Y, Li P, Liu YD, Yao ZP, Xin GZ, et al. Untargeted metabolomics coupled with chemometric analysis reveals species-specific steroidal alkaloids for the authentication of medicinal fritillariae bulbus and relevant products. *J Chromatogr A* 2020;**1612**:460630.
150. Zhang W, Jiang H, Yang J, Song G, Wen D, Liu W, et al. A high-throughput metabolomics approach for the comprehensive differentiation of four pulsatilla adans herbs combined with a nontargeted bidirectional screen for rapid identification of triterpenoid saponins. *Anal Bioanal Chem* 2019;**411**:2071−88.
151. Yao R, Heinrich M, Zou Y, Reich E, Zhang X, Chen Y, et al. Quality variation of goji (fruits of lycium spp.) in China: a comparative morphological and metabolomic analysis. *Front Pharmacol* 2018;**9**:151.
152. Nam M, Jo SR, Kim YC, Kim MS. UPLC−QTOF-MS-based metabolomics and antioxidant capacity of codonopsis lanceolata from different geographical origins. *Foods* 2023;**12**:267.
153. Zhang S, Li C, Gu W, Qiu R, Chao J, Pei L, et al. Metabolomics analysis of dandelions from different geographical regions in China. *Phytochem Anal* 2021;**32**:899−906.
154. Wu D, Yu D, Zhang Y, Dong J, Li D, Wang D. Metabolite profiles, bioactivity, and HPLC fingerprint of different varieties of eucommia ulmoides oliv.: towards the utilization of medicinal and commercial chinese endemic tree. *Molecules* 2018;**23**:1898.
155. Liu PP, Shan GS, Zhang F, Chen JN, Jia TZ. Metabolomics analysis and rapid identification of changes in chemical ingredients in crude and processed astragali radix by UPLC−QTOF-MS combined with novel informatics UNIFI platform. *Chin J Nat Med* 2018;**16**:714−20.
156. Shin JH, Park YJ, Kim W, Kim DO, Kim BY, Lee H, et al. Change of ginsenoside profiles in processed ginseng by drying, steaming, and puffing. *J Microbiol Biotechnol* 2019;**29**:222−9.
157. Lai JL, Deng ZX, Ji XH, Luo XG. Absorption and interaction mechanisms of uranium & cadmium in purple sweet potato (*Ipomoea batatas* L.). *J Hazard Mater* 2020;**400**:123264.
158. Mattoli L, Gianni M, Burico M. Mass spectrometry-based metabolomic analysis as a tool for quality control of natural complex products. *Mass Spectrom Rev* 2023;**42**:1358−96.
159. Tang C, Zhou Y, Zhang R, Wei P, Xie P, Liao X. Metabolomics application in quality evaluation of chinese medicine and research on mechanism ofantibacterial and anti-inflammatory effects. *China Mod Tradit Chin Med* 2021;**23**:1664−70.
160. Xiao Q, Mu X, Liu J, Li B, Liu H, Zhang B, et al. Plant metabolomics: a new strategy and tool for quality evaluation of chinese medicinal materials. *Chin Med* 2022;**17**:45.
161. van Dijk ADJ, Kootstra G, Kruijer W, de Ridder D. Machine learning in plant science and plant breeding. *iScience* 2021;**24**:101890.
162. Farooq MA, Gao S, Hassan MA, Huang Z, Rasheed A, Hearne S, et al. Artificial intelligence in plant breeding. *Trends Genet* 2024;**40**:891−908.
163. Armstrong EM, Larson ER, Harper H, Webb CR, Dohleman F, Araya Y, et al. One hundred important questions facing plant science: an international perspective. *New Phytol* 2023;**238**:470−81.
164. Juliana P, Poland J, Huerta-Espino J, Shrestha S, Crossa J, Crespo-Herrera L, et al. Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat Genet* 2019;**51**:1530−9.
165. Uygun S, Azodi CB, Shiu SH. *Cis*-regulatory code for predicting plant cell-type transcriptional response to high salinity. *Plant Physiol* 2019;**181**:1739−51.

166. Sartor RC, Noshay J, Springer NM, Briggs SP. Identification of the expressome by machine learning on omics data. *Proc Natl Acad Sci U S A* 2019;**116**:18119−25.
167. Zheng Z, Li S, Su J, Leung AW, Lam TW, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci* 2022;**2**:797−803.
168. Fei S, Hassan MA, Xiao Y, Rasheed A, Xia X, Ma Y, et al. Application of multi-layer neural network and hyperspectral reflectance in genome-wide association study for grain yield in bread wheat. *Field Crops Res* 2022;**289**:108730.
169. Montesinos-López A, Crespo-Herrera L, Dreisigacker S, Gerard G, Vitale P, Saint Pierre C, et al. Deep learning methods improve genomic prediction of wheat breeding. *Front Plant Sci* 2024;**15**: 1324090.
170. Varghese R, Shringi H, Efferth T, Ramamoorthy S. Artificial intelligence driven approaches in phytochemical research: trends and prospects. *Phytochem Rev* 2025;**24**:3649−64.
171. Pan Y, Zhang H, Chen Y, Gong X, Yan J, Zhang H. Applications of hyperspectral imaging technology combined with machine learning in quality control of traditional chinese medicine from the perspective of artificial intelligence: a review. *Crit Rev Anal Chem* 2024;**54**: 2850−64.
172. Varghese R, Cherukuri AK, Doddrell NH, Doss CGP, Simkin AJ, Ramamoorthy S. Machine learning in photosynthesis: prospects on sustainable crop development. *Plant Sci* 2023;**335**:111795.
173. Saha D, Manickavasagan A. Machine learning techniques for analysis of hyperspectral images to determine quality of food products: a review. *Curr Res Food Sci* 2021;**4**:28−44.
174. Tomar M, Bhardwaj R, Kumar M, Singh SP, Krishnan V, Kansal R, et al. Development of nir spectroscopy based prediction models for nutritional profiling of pearl millet (*Pennisetum glaucum* (L.)) R. Br: a chemometrics approach. *LWT (Lebensm-Wiss & Technol)* 2021; **149**:111813.
175. Liu L, Qu H. Recent advancement of chemical imaging in pharmaceutical quality control: from final product testing to industrial utilization. *J Innovative Opt Health Sci* 2020;**13**:1930014.
176. Wang C, Liu B, Liu L, Zhu Y, Hou J, Liu P, et al. A review of deep learning used in the hyperspectral image analysis for agriculture. *Artif Intell Rev* 2021;**54**:5205−53.
177. Hernandes IS, Da Silva HC, Dos Santos HF, De Almeida WB. Conformational analysis of 5,4′-dihydroxy-7,5′,3′-trimethoxyisoflavone in solution using ^{1}H NMR: a density functional theory approach. *J Phys Chem A* 2020;**124**:5182−93.
178. Arya S, Sandhu KS, Singh J, kumar S. Deep learning: as the new frontier in high-throughput plant phenotyping. *Euphytica* 2022;**218**: 47.
179. Das Choudhury S, Samal A, Awada T. Leveraging image analysis for high-throughput plant phenotyping. *Front Plant Sci* 2019;**10**:508.
180. Zhou L, Zhang C, Taha MF, Wei X, He Y, Qiu Z, et al. Wheat kernel variety identification based on a large near-infrared spectral dataset and a novel deep learning-based feature selection method. *Front Plant Sci* 2020;**11**:575810.
181. Sandhu KS, Lozada DN, Zhang Z, Pumphrey MO, Carter AH. Deep learning for predicting complex traits in spring wheat breeding program. *Front Plant Sci* 2020;**11**:613325.
182. Zhu W, Sun Z, Yang T, Li J, Peng J, Zhu K, et al. Estimating leaf chlorophyll content of crops *via* optimal unmanned aerial vehicle hyperspectral data at multi-scales. *Comput Electron Agric* 2020;**178**: 105786.
183. Esgario JGM, Krohling RA, Ventura JA. Deep learning for classification and severity estimation of coffee leaf biotic stress. *Comput Electron Agric* 2020;**169**:105162.
184. Waldmann P, Pfeiffer C, Mészáros G. Sparse convolutional neural networks for genome-wide prediction. *Front Genet* 2020;**11**: 25.
185. Gao F, Fu L, Zhang X, Majeed Y, Li R, Karkee M, et al. Multi-class fruit-on-plant detection for apple in snap system using faster R-CNN. *Comput Electron Agric* 2020;**176**:105634.
186. Malik OA, Ismail N, Hussein BR, Yahya U. Automated real-time identification of medicinal plants species in natural environment using deep learning models-A case study from borneo region. *Plants* 2022;**11**:1952.
187. Ni P, Huang N, Nie F, Zhang J, Zhang Z, Wu B, et al. Genome-wide detection of cytosine methylations in plant from nanopore data using deep learning. *Nat Commun* 2021;**12**:5976.
188. Pandita D, Pandita A, Wani SH, Abdelmohsen SAM, Alyousef HA, Abdelbacki AMM, et al. Crosstalk of multi-omics platforms with plants of therapeutic importance. *Cells* 2021;**10**:1296.
189. Zhang W, Zeng Y, Jiao M, Ye C, Li Y, Liu C, et al. Integration of high-throughput omics technologies in medicinal plant research: the new era of natural drug discovery. *Front Plant Sci* 2023;**14**: 1073848.
190. Yang L, Yang Y, Huang L, Cui X, Liu Y. From single- to multi-omics: future research trends in medicinal plants. *Briefings Bioinf* 2023;**24**:bbac485.
191. Weckwerth W, Ghatak A, Bellaire A, Chaturvedi P, Varshney RK. Panomics meets germplasm. *Plant Biotechnol J* 2020;**18**:1507−25.
192. Zhou M, Yang G, Sun G, Guo Z, Gong X, Pan Y. Resolving complicated relationships of the *Panax bipinnatifidus* complex in southwestern China by RAD-seq data. *Mol Phylogenet Evol* 2020; **149**:106851.
193. Luo M, Li A, Wang F, Jiang J, Wang Z, You J. Integrative analysis of multiple metabolomes and transcriptome revealed color expression mechanism in red skin root syndrome of *Panax ginseng*. *Ind Crops Prod* 2022;**177**:114491.
194. Cui Y, Fan B, Xu X, Sheng S, Xu Y, Wang X. A high-density genetic map enables genome synteny and QTL mapping of vegetative growth and leaf traits in gardenia. *Front Genet* 2021;**12**:802738.
195. Zhang Y, Cui J, Hu H, Xue J, Yang J, Xu J. Integrated four comparative-omics reveals the mechanism of the terpenoid biosynthesis in two different overwintering cryptomeria fortunei phenotypes. *Front Plant Sci* 2021;**12**:740755.
196. Peng Z, Wang Y, Zuo WT, Gao YR, Li RZ, Yu CX, et al. Integration of metabolome and transcriptome studies reveals flavonoids, abscisic acid, and nitric oxide comodulating the freezing tolerance in liriope spicata. *Front Plant Sci* 2021;**12**:764625.
197. Li C, Wood JC, Vu AH, Hamilton JP, Rodriguez Lopez CE, Payne RME, et al. Single-cell multi-omics in the medicinal plant *Catharanthus roseus*. *Nat Chem Biol* 2023;**19**:1031−41.
198. Tran TP, Din FU, Brankovic L, Sanin C, Hester SM. A systematic review of medicinal plant identification using deep learning. In: *Asian Conference on Intelligent Information and Database Systems*. **14796**. Singapore, ACIIDS: Springer; 2024.
199. Mulugeta AK, Sharma DP, Mesfin AH. Deep learning for medicinal plant species classification and recognition: a systematic review. *Front Plant Sci* 2023;**14**:1286088.
200. Chanyal H, Yadav RK, Saini DKJ. Classification of medicinal plants leaves using deep learning technique: a review. *Int J Intell Syst Appl Eng* 2022;**10**:78−87.
201. Roy SK, Krishna G, Dubey SR, Chaudhuri BB. Hybridsn: exploring 3-D−2-D CNN feature hierarchy for hyperspectral image classification. *IEEE Geosci Remote Sens Lett* 2020;**17**:277−81.
202. Ramesh R, Sankaradass V, Nishintha A, HarshivKrishna M, Ponnusamy R. Applying vision transformers for herbal medicine classification:a novel approach to plant identification. In 2025 international conference on data science, agents & artificial intelligence (ICDSAAI). Available from: http://doi: 10.1109/ICDSAAI65575.2025.11011692.
203. Architha SP, Patil AP. *Classification of medicinal leaves using support vector machine, convolutional neural network and you only look once. 2020 IEEE Bangalore humanitarian technology conference*. B-HTC). Vijiyapur, India; 2020. p. 1−6. Available from: https://doi.org/10.1109/B-HTC50970.2020.9297878.
204. Abdollahi J. Identification of medicinal plants in ardabil using deep learning: identification of medicinal plants using deep learning. In: *2022 27th international computer conference, computer Society of*

Iran (CSICC); 2022. p. 1−6. Available from: https://doi.org/10.1109/CSICC55295.2022.9780493. Tehran, Iran.
205. Kumar P, Vinod KS, Senthil PS. CNN and edge-based segmentation for the identification of medicinal plants. In: *2024 5th international conference on intelligent communication technologies and virtual Mobile networks (ICICV)*. India: Tirunelveli; 2024. p. 89−94. Available from: https://doi.org/10.1109/ICICV62344.2024.00021.
206. Xu M, Kim H, Yang J, Fuentes A, Meng Y, Yoon S, et al. Embracing limited and imperfect training datasets: opportunities and challenges in plant disease recognition using deep learning. *Front Plant Sci* 2023;**14**:1225409.
207. Pushpa BR, Shree Hari AV. AshokA. Diseased leaf segmentation from complex background using indices based histogram. In: *2021 6th international conference on communication and electronics systems (ICCES)*. India: Coimbatre; 2021. p. 1502−7. Available from: https://doi.org/10.1109/ICCES51350.2021.9489112.
208. Azadnia R, Noei-Khodabadi F, Moloudzadeh A, Jahanbakhshi A, Omid M. Medicinal and poisonous plants classification from visual characteristics of leaves using computer vision and deep neural networks. *Ecol Inform* 2024;**82**:102683.
209. Iram A, Dong Y, Ignea C. Synthetic biology advances towards a bio-based society in the era of artificial intelligence. *Curr Opin Biotechnol* 2024;**87**:103143.
210. Reinhardt JK, Craft D, Weng JK. Toward an integrated omics approach for plant biosynthetic pathway discovery in the age of AI. *Trends Biochem Sci* 2025;**50**:311−21.
211. Ignea C, Pontini M, Motawia MS, Maffei ME, Makris AM, Kampranis SC. Synthesis of 11-carbon terpenoids in yeast using protein and metabolic engineering. *Nat Chem Biol* 2018;**14**:1090−8.
212. Herbert AJ, Shepherd SA, Cronin VA, Bennett MR, Sung R, Micklefield J. Engineering orthogonal methyltransferases to create alternative bioalkylation pathways. *Angew Chem Int Ed Engl* 2020;**59**:14950−6.
213. Ignea C, Raadam MH, Motawia MS, Makris AM, Vickers CE, Kampranis SC. Orthogonal monoterpenoid biosynthesis in yeast constructed on an isomeric substrate. *Nat Commun* 2019;**10**:3799.
214. Wen Z, Zhang ZM, Zhong L, Fan J, Li M, Ma Y, et al. Directed evolution of a plant glycosyltransferase for chemo- and regioselective glycosylation of pharmaceutically significant flavonoids. *ACS Catal* 2021;**11**:14781−90.
215. Gelder KV, Oliveira-Filho ER, García-García JD, Hu Y, Bruner SD, Hanson AD. Directed evolution of aerotolerance in sulfide-dependent thiazole synthases. *ACS Synth Biol* 2023;**12**:963−70.
216. Luo X, Reiter MA, d'Espaux L, Wong J, Denby CM, Lechner A, et al. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* 2019;**567**:123−6.
217. Zhang J, Hansen LG, Gudich O, Viehrig K, Lassen LMM, Schrübbers L, et al. A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* 2022;**609**:341−7.
218. Reed J, Orme A, El-Demerdash A, Owen C, Martin LBB, Misra RC, et al. Elucidation of the pathway for biosynthesis of saponin adjuvants from the soapbark tree. *Science* 2023;**379**:1252−64.
219. Jiang B, Gao L, Wang H, Sun Y, Zhang X, Ke H, et al. Characterization and heterologous reconstitution of taxus biosynthetic enzymes leading to baccatin iii. *Science* 2024;**383**:622−9.
220. Ignea C, Raadam MH, Koutsaviti A, Zhao Y, Duan YT, Harizani M, et al. Expanding the terpene biosynthetic code with non-canonical 16 carbon atom building blocks. *Nat Commun* 2022;**13**:5188.
221. Zhao M, Che Y, Gao Y, Zhang X. Application of multi-omics in the study of traditional chinese medicine. *Front Pharmacol* 2024;**15**:1431862.
222. Li S, Zhang B. Traditional chinese medicine network pharmacology: theory, methodology and application. *Chin J Nat Med* 2013;**11**:110−20.
223. Li S. Mapping ancient remedies: applying a network approach to traditional chinese medicine. *Science* 2015;**350**:S72−4.
224. Wang B, Zhang T, Liu Q, Sutcharitchan C, Zhou Z, Zhang D, et al. Elucidating the role of artificial intelligence in drug development from the perspective of drug−target interactions. *J Pharm Anal* 2025;**15**:101144.
225. Fan Q, Wu C, Du Y, Wang B, Xie Y, Zhang Z, et al. Comparison of jinzhen oral liquid and ambroxol hydrochloride and clenbuterol hydrochloride oral solution in the treatment of acute bronchitis in children: a multicenter, non-inferiority, prospective, randomized controlled trial. *Acta Pharm Sin B* 2024;**14**:5186−200.
226. Li Y, Wang X, Yang Z, Liao NX, Wang YW, Ji CF, et al. A new approach for optimizing empirical prescriptions of famous physicians based on network target: taking qingluo decoction as an example. *J Chin Materia Medica* 2022;**47**:5264−73.
227. Wang B, Kong D, Yang Z, Kang J, Sun D, Duan X, et al. Revealing the mechanisms of compound kushen injection on oxidative stress regulation in the treatment of radiation-induced lung injury. *Engineering* 2025. Available from: https://doi.org/10.1016/j.eng.2025.09.018.
228. Wang Z, Zhang T, Wang B, Li S. TCM network pharmacology: new perspective integrating network target with artificial intelligence and multi-modal multi-omics technologies. *Chin J Nat Med* 2025;**23**:1425−34.
229. Zhang P, Zhang D, Zhou W, Wang L, Wang B, Zhang T, et al. Network pharmacology: towards the artificial intelligence-based precision traditional chinese medicine. *Briefings Bioinf* 2023;**25**:bbad518.
230. Zhao W, Wang B, Li S. Network pharmacology for traditional chinese medicine in era of artificial intelligence. *Chin Herb Med* 2024;**16**:558−60.
231. Duan P, Yang K, Su X, Fan S, Dong X, Zhang F, et al. Htinet2: herb−target prediction *via* knowledge graph embedding and residual-like graph neural network. *Briefings Bioinf* 2024;**25**:bbae414.
232. Li P, Zhang H, Zhang W, Zhang Y, Zhan L, Wang N, et al. Tmnp: a transcriptome-based multi-scale network pharmacology platform for herbal medicine. *Briefings Bioinf* 2022;**23**:bbab542.
233. Wang B, Pan B, Zhang T, Liu Q, Li S. Transfer learning and permutation-invariance improving predicting genome-wide, cell-specific and directional interventions effects of complex systems. *Adv Sci* 2025;**12**:e09456.
234. Guo R, Luo X, Liu J, Liu L, Wang X, Lu H. Omics strategies decipher therapeutic discoveries of traditional chinese medicine against different diseases at multiple layers molecular-level. *Pharmacol Res* 2020;**152**:104627.
235. Wang B, Xiao L, Chen P, Zhang T, Zhang P, Cao L, et al. Uncovering the role of traditional chinese medicine in immune-metabolic balance of gastritis from the perspective of cold and hot: Jin hong tablets as a case study. *Chin Med* 2024;**19**:134.
236. Wang B, Zhou W, Zhang H, Wang W, Zhang B, Li S. Exploring the effect of weifuchun capsule on the toll-like receptor pathway mediated hes6 and immune regulation against chronic atrophic gastritis. *J Ethnopharmacol* 2023;**303**:115930.
237. Chandrasekhar V, Rajan K, Kanakam SRS, Sharma N, Weißenborn V, Schaub J, et al. Coconut 2.0: a comprehensive overhaul and curation of the collection of open natural products database. *Nucleic Acids Res* 2025;**53**:D634−43.
238. Santiago-de-la-Cruz JA, Rivero-Segura NA, Gomez-Verjan JC. Structure-based machine learning screening identifies natural product candidates as potential geroprotectors. *J Cheminf* 2025;**17**:106.
239. Norouzi R, Norouzi R, Abbasi K, Norouzi R, Razzaghi P. Dft_anpd: a dual-feature two-sided attention network for anticancer natural products detection. *Comput Biol Med* 2025;**194**:110442.
240. Mangal M, Sagar P, Singh H, Raghava GP, Agarwal SM. Npact: naturally occurring plant-based anti-cancer compound-activity-target database. *Nucleic Acids Res* 2013;**41**:D1124−9.
241. Zhao H, Yang Y, Wang S, Yang X, Zhou K, Xu C, et al. Npass database update 2023: quantitative natural product activity and species source database for biomedical research. *Nucleic Acids Res* 2023;**51**:D621−8.
242. Yang Y, Rao Y, Yu M, Kang Y. Multi-layer information fusion based on graph convolutional network for knowledge-driven herb recommendation. *Neural Netw* 2022;**146**:1−10.

243. Ru J, Li P, Wang J, Zhou W, Li B, Huang C, et al. Tcmsp: a database of systems pharmacology for drug discovery from herbal medicines. *J Cheminf* 2014;**6**:13.
244. Zhang Y, Li X, Shi Y, Chen T, Xu Z, Wang P, et al. Etcm v2.0: an update with comprehensive resource and rich annotations for traditional chinese medicine. *Acta Pharm Sin B* 2023;**13**:2559−71.
245. Wishart DS, Sajed T, Pin M, Poynton EF, Goel B, Lee BL, et al. The natural products magnetic resonance database (NP-MRD) for 2025. *Nucleic Acids Res* 2024;**53**:D700−8.
246. Chen W, Yu Z, Leng L, Sun D, Liu H, Gong RZ, et al. Artificial intelligence-curated repository of gene-encoded natural diverse components from herbal medicines. *Innovation* 2025;**6**:101011.
247. Hou D, Lin H, Feng Y, Zhou K, Li X, Yang Y, et al. Cmaup database update 2024: extended functional and association information of useful plants for biomedical research. *Nucleic Acids Res* 2024;**52**: D1508−18.
248. Poynton EF, van Santen JA, Pin M, Contreras MM, McMann E, Parra J, et al. The natural products atlas 3.0: extending the database of microbially derived natural products. *Nucleic Acids Res* 2025;**53**: D691−9.
249. Harrington RA, Adhikari V, Rayner M, Scarborough P. Nutrient composition databases in the age of big data: Fooddb, a comprehensive, real-time database infrastructure. *BMJ Open* 2019;**9**: e026652.
250. Lyu C, Chen T, Qiang B, Liu N, Wang H, Zhang L, et al. Cmnpd: a comprehensive marine natural products database towards facilitating drug discovery from the ocean. *Nucleic Acids Res* 2021;**49**: D509−15.
251. Wang Q, Luo S, Yang Y, Bai Y, Wei J, Xu KW, et al. Wp-mod: a multi-omics and taxonomy database for woody plants. *Plant Commun* 2025;**6**:101290.
252. Hawkins C, Xue B, Yasmin F, Wyatt G, Zerbe P, Rhee Seung Y. Plant metabolic network 16: expansion of underrepresented plant groups and experimentally supported enzyme data. *Nucleic Acids Res* 2024; **53**:D1606−13.
253. Valentin G, Abdel T, Gaëtan D, Jean-François D, Matthieu C, Mathieu R. Greenphyldb v5: a comparative pangenomic database for plant genomes. *Nucleic Acids Res* 2021;**49**:D1464−71.
254. Su X, Yang L, Wang D, Shu Z, Yang Y, Chen S, et al. 1 K medicinal plant genome database: an integrated database combining genomes and metabolites of medicinal plants. *Hortic Res* 2022;**9**:uhac075.
255. Chen H, Yin X, Guo L, Yao J, Ding Y, Xu X, et al. Plantscrnadb: a database for plant single-cell rna analysis. *Mol Plant* 2021;**14**:855−7.
256. Meng F, Tang Q, Chu T, Li X, Lin Y, Song X, et al. Tcmpg: an integrative database for traditional chinese medicine plant genomes. *Hortic Res* 2022;**9**:uhac060.
257. Gao K, Liu L, Lei S, Li Z, Huo P, Wang Z, et al. Herb 2.0: an updated database integrating clinical and experimental evidence for traditional chinese medicine. *Nucleic Acids Res* 2025;**53**:D1404−14.
258. Lv Q, Chen G, He H, Yang Z, Zhao L, Zhang K, et al. TCMbank-the largest TCM database provides deep learning-based Chinese-Western medicine exclusion prediction. *Signal Transduct Targeted Ther* 2023; **8**:127.
259. Hasselgren C, Oprea TI. Artificial intelligence for drug discovery: are we there yet?. *Annu Rev Pharmacol Toxicol* 2024;**64**:527−50.
260. Cerchia C, Lavecchia A. New avenues in artificial-intelligence-assisted drug discovery. *Drug Discov Today* 2023;**28**:103516.
261. Gangwal A, Lavecchia A. Unleashing the power of generative ai in drug discovery. *Drug Discov Today* 2024;**29**:103992.
262. Gangwal A, Lavecchia A. Artificial intelligence in natural product drug discovery: current applications and future perspectives. *J Med Chem* 2025;**68**:3948−69.
263. Wolfender JL, Litaudon M, Touboul D, Queiroz EF. Innovative omics-based approaches for prioritisation and targeted isolation of natural products−new strategies for drug discovery. *Nat Prod Rep* 2019;**36**:855−68.