Natural Resources for Human Health

visagaa publishing house

Review

View Article online

Check for updates

Received 29 October 2025
Revised 21 November 2025
Accepted 25 January 2026
Available Online 30 June 2026

Edited by Kannan RR Rengasamy

KEYWORDS:
Artificial Intelligence
Machine Learning
Pharmacokinetics
Multi-omics
Natural Products
Metabolomics

https://doi.org/10.53365/nrfhh/217344
eISSN: 2583-1194

Integrating Artificial Intelligence and Machine Learning in Natural Product Discovery: From Omics Data to Drug Design

J. Geetha[1], Nathiya Ranganathan[1], G. Gulothungan[2], Hitesh Chopra[3,*]

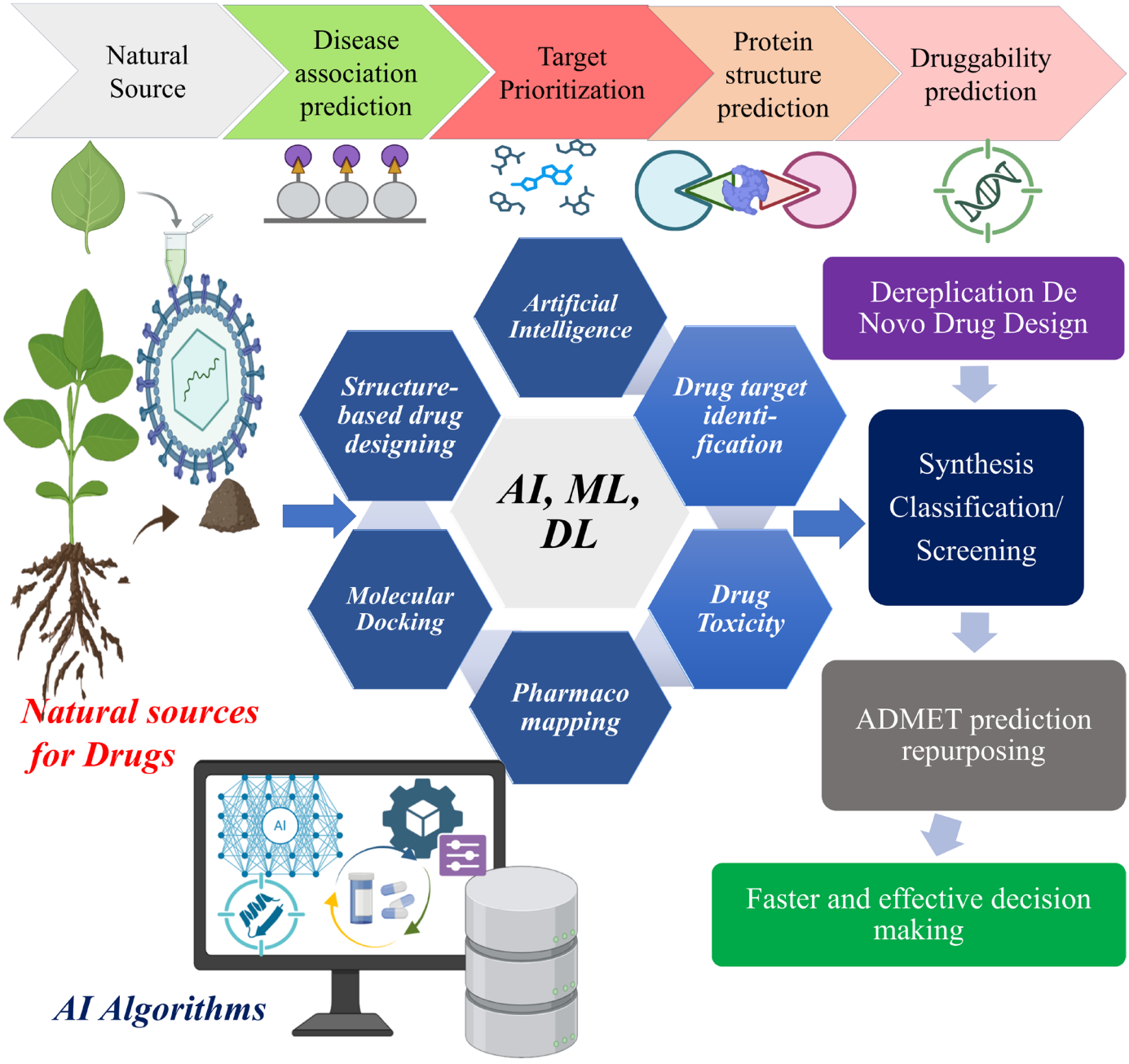
[1]Department of Microbiology, Auxilium College (Autonomous), Vellore, India
[2]Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, India
[3]Chitkara College of Pharmacy, Center for Research Impact & Outcome, Chitkara University, Rajpura, Punjab, India

ABSTRACT: Natural products represent an essential source of bioactive materials with potent biological activity, traditional approaches for their discovery and development remain time consuming and resource-intensive. Recent advances of Artificial Intelligence (AI) and machine learning (ML) are transforming natural product research by enabling predictive modeling, high throughput screening, automated compound identification and integration of complex omics datasets. Other AI and ML methods such as deep learning, neural networks and generative algorithms are enhancing drug discovery by predicting bioactivity and structural elucidation, and optimizing pharmacokinetic properties of natural products. Furthermore, AI, ML, and other technologies are opening new possibilities, such as genome mining for biosynthetic gene clusters, promoting dereplication strategies, and integrating multi-omics datasets to understand natural product biosynthesis and pharmacology at the systems level. This review summarizes the recent applications of AI and ML in natural product discovery, metabolomics, cheminformatics and pharmacology, present highlighted successful case studies, associated challenges and future perspectives for AI-driven sustainable drug discovery.

* Corresponding author.
E-mail address: chopraontheride@gmail.com (Hitesh Chopra)

GRAPHICAL ABSTRACT

Integrating Artificial Intelligence And Machine Learning In Natural Product Discovery



1. INTRODUCTION

Natural products are bioactive agents derived from plants, animals, fungi, and microorganisms, and are more commonly referred to as secondary metabolites because though they are excluded in essential life processes, such as in growth, reproduction, or development, they play an important role in ecological interactions (defense, competition, and communication). For thousands of years, natural products have acted as a primary source of food, traditional medicines, and health-promoting agents, and formed the basis of ancient medical systems (Ayurveda, Traditional Chinese Medicine, and Unani). In drug discovery, natural products provide unparalleled molecular diversity and a range of biological

activities, making them an important source for therapeutic leads. For example, penicillin was obtained from the fungus *Penicillium*, the first successful antibiotic used for treatment, and lovastatin was obtained from *Aspergillus terreus*, a precursor in statin cholesterol-lowering medications (Dias et al., 2012; Thomford et al., 2018; Newman et al., 2020; Atanasov et al., 2021).

However, there are considerable barriers to discovering new drugs among natural products. Identifying bioactive molecules is a time-consuming and labor-intensive process consisting of extraction, isolation, and characterization. These protocols utilize bioassay-guided fractionation, solvent extraction, and chromatographic separation, culminating in structural elucidation by advanced spectroscopy techniques such as Nuclear Magnetic Resonance (NMR), Mass Spectrometry (MS), or X-ray crystallography. Also, the potential rediscovery of the same compound (dereplication), low yields, and complicated purification steps can limit throughput and scalability (Hou et al., 2020).

Conventional research in natural products faces a range of scientific, methodological, and environmental barriers. Many natural products are unstable (due to their complex structures) or difficult to synthesize and cannot be scaled up in-line with their complexity. Another issue is the variation in concentration of active compounds when testing crude extracts, which can impact the chemical composition analysis and the reproducibility of results during analysis. These can also cause environmental issues, as many extraction processes consume large volumes of organic solvents and contribute to pollution and energy wastage, as well as ecological impact through overextraction of particular natural products. The lack of selectivity for concurrent impurities culminates in the reduced efficiency of many bioassays as the screening process is slowed due to impurities. As a result, conventional natural product research is much slower, less efficient, often more resource intensive, and environmentally detrimental compared to modern approaches that allow for more systematic research through the use of new technologies (Simoben et al., 2023; Mungwari et al., 2024).

Artificial Intelligence (AI) and Machine Learning (ML) are making significant changes in natural product research, enhancing discover, improving efficiency and presenting new tools for traditionally slow and labor-intensive processes. These technologies have dramatically increased the pace of screening large chemicals and accurately predict their pharmacological properties, supporting the discovery and identification of bioactive compounds. AI-based genome mining reveals new natural products present in plants and microbes that were unidentified before, and dereplication tools help not to isolate their natural products again, saving time and money. Deep learning (DL) and ML algorithms are utilized to predict structure and biological activity of the compounds, and identify potential drug targets for lead optimization. AI/ML models are also being employed to improve and optimize the bioactive compound production in microbes. Target identification has also significantly improved, as AI tools can "deorphanize" compounds by predicting their protein targets and elucidating the mechanism of action. In general, these technologies have greatly advanced how quickly and efficiently natural products can be discovered, handled, and analyzed, aided in overcoming longstanding technical and analytical challenges in natural product research to open pathways, and proposed setting for new drug discovery and development (Sarkar et al., 2023; Gangwal et al., 2025; Fu et al., 2025).

This review aims to investigate the recent advances in the application of AI and ML in natural product research and drug discovery. It focuses on AI and ML-based predictive modeling, compound identification, dereplication, genome mining of biosynthetic gene clusters, and integration of multi-omics datasets for understanding natural product biosynthesis and pharmacology. Relevant peer-reviewed articles published between 2011 and 2025 were collected from databases including PubMed, ScienceDirect, Scopus, and Google Scholar, and from journals published by MDPI, Elsevier, Springer Nature, Wiley, and Frontiers. Search keywords included natural products, AI, ML, cheminformatics, metabolomics, and drug discovery. The articles were selected on the relevance of natural product studies utilizing AI/ML techniques to provide a summary of the latest trends, existing challenges, and future directions in sustainable discovery of drugs.

2. ROLE OF AI AND MACHINE LEARNING IN NATURAL PRODUCT RESEARCH

Artificial Intelligence (AI) is designated as machine-based systems that has the ability to predict, recommend, or make a decision based on human-specified objectives with consequences that can affect real or virtual environments. AI systems observe inputs from machines and humans, transform the observations into models, and then infer from those models to create actionable options. A key subset of AI in drug development is ML, in which an algorithm is trained to improve its performance using data (https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug, accessed on 22/10/2025).

Artificial Intelligence has fundamentally reshaped drug development in all phases of the drug's life cycle, including nonclinical, clinical, post marketing and manufacturing. The expansion of AI is evident in the increasing number of

drug applications submitted to the FDA's Center for Drug Evaluation and Research (CDER) that utilize AI. In addition to discovery and evaluation, AI is critical for enabling digital health technologies (DHTs) and advancing the analysis of real-world data to facilitate more precise, impactful2, and data-driven patient care and therapy development. CDER is committed to integrating AI responsibly by employing a risk-based regulatory framework to balance innovation with patient safety.

Artificial Intelligence in drug discovery uses ML algorithms for large biological datasets, in processing, predicting results, identifying targets, and optimizing the drug. ML algorithms learn from data to recognize patterns and make predictions. AI systems provide data-driven insights by discovering hidden patterns in genomics, proteomics, and patient data, enabling predictive modeling of bioactivity, toxicity, and pharmacokinetic properties. In addition, ML uses iterative predictive space learning to continuously optimize compounds. AI also supports hypothesis formulation that drives experimental research and provides hypotheses that allow us to rationally tailor treatments to the individual, such as lifestyle and genetics, ultimately improving efficacy and minimizing side effects (Paul et al., 2021; Moingeon et al., 2022; Vora et al., 2023)

Artificial Intelligence and ML have gained considerable status in the discovery of new drugs from natural products in the past few years. The main algorithms used are DL, which includes neural networks (e.g., Convolutional Neural Networks [CNNs] and Graph Neural Networks [GNNs]), Random Forest, and Support Vector Machines (SVMs). These computational methods are applied to different tasks of prediction of molecular properties, virtual screening of drug candidates, classification of compounds, and prediction of drug–target interactions. The researchers can utilize these algorithms to accelerate bioactive molecules, develop pharmacokinetic and pharmacodynamic properties, and reduce the time and expense of the traditional drug discovery (Table 1).

2.1. Integration of AI in natural product databases and informatics

Incorporating AI into natural product databases and informatics has transformed drug discovery through the automation of laborious methodologies, increasing throughput and assisting in the analysis and interpretation of large and complex datasets and significant time savings and efficiency. Algorithms can predict compound characteristics, propose new structures, integrate genomic and metabolomic data, and assist in synthetic route planning to improve the accuracy of bioactivity predictions. AI can rapidly search, classify, and dereplicate complex compounds and enhance natural product research while minimizing experimental errors and providing greater predictive accuracy for toxicity and efficacy. It allows the integration of multiple datasets and uncovers concealed patterns of new compounds that are useful for understanding the synthetic pathway of complex natural products. Predictive modeling allows the prioritization of most promising compounds by predicting their molecular properties, biological activities, and protein interactions. Some of its challenges are as follows: the dependence on good quality and comprehensive data, the complexities of integrating different datasets, transparency and reproducibility, concerns over ethical issues of data use, and the need to involve experts, scientists, and AI researchers from different disciplines. Overall, AI-driven informatics improve the natural product research by emphasizing careful data management and validation (Blanco-González et al., 2023; Sathya et al., 2025)

Table 1
Key AI/ML techniques (deep learning/neural network algorithms) used in natural product research for drug discovery.

Algorithm	Description	Application in drug discovery	References
Convolutional Neural Networks (CNNs)	Analyze structured data such as images or molecular grids	Predicts molecular properties, activity modeling, image-based screening	(Dara et al., 2022; Obaido et al., 2024; Patel et al., 2020)
Graph Neural Networks (GNNs)	Learn latent features of molecules represented as graphs	Predict drug response, model drug–target interactions, understand molecular mechanisms	(Dara et al., 2022; Obaido et al., 2024; Patel et al., 2020)
Generative Adversarial Networks (GANs)	Generative AI framework with two competing networks	Design novel molecular structures, create analogs of natural products	(Dara et al., 2022; Obaido et al., 2024; Patel et al., 2020)
Random Forest	A highly robust, supervised learning technique that integrates multiple decision to improve prediction accuracy	Feature selection, compound classification, toxicity prediction, virtual screening of large compound libraries	(Dara et al., 2022; Obaido et al., 2024; Patel et al., 2020)
Support Vector Machines (SVM)	Supervised an algorithm that separates classes by finding an optimal hyperplane	Classifying the active and inactive compounds, ranking potential drug candidates, predicting drug–target interactions by integrating ligand and protein features	(Dara et al., 2022; Obaido et al., 2024; Patel et al., 2020; Singh et al., 2023)

3. APPLICATIONS OF AI AND ML IN NATURAL PRODUCT DISCOVERY

The use of AI and ML in drug development has changed natural product discovery by automating some of the essential steps of the drug development. From predicting the biological activity to predicting toxicity, identifying molecular structures and optimizing lead compounds, AI-based tools improve research workflows and enhance decision-making. Figure 1 depicts the major applications of AI and ML in natural product discovery, including predictive modeling, automated identification of compound, and high-throughput virtual screening.

3.1. Predictive modeling and bioactivity prediction

Predictive modeling and bioactivity prediction are approaches that leverage AI and ML to rapidly connect molecular structures with their biological effects. By predicting a compound's efficacy, toxicity, and mechanism of action, these methods allow the researchers to efficiently prioritize promising drug candidates and significantly reduce necessary experimental work.

Predicting pharmacological activity from molecular structures depends on computational methods, notably virtual screening and Quantitative Structure–Activity Relationship (QSAR) modeling, both of which have been greatly enhanced by the use of AI. Virtual screening allows for the efficient identification of large libraries, in silico, for downstream laboratory testing based on predicting similar molecules to known actives (ligand-based) or predicting interacting similarities based on predicted structure of 3D target (structure-based). AI makes a substantial contribution to this process, enabling high-speed evaluations of massive amounts of data, as well as an improvement in the accuracy of estimates. QSAR modeling predicts a compound's biological activity from its chemical structure using statical or ML methods. These predictive models allow accurate prediction of activity for new molecules based on existing compounds with known activity. AI-driven QSAR models, including DL models, allow for substantially more powerful and dependable predictions based on molecular descriptors. They provide significant advantages by improving the speed of drug discovery, reducing the experimental costs, and facilitating predictive properties such as activity, toxicity, and resistance. However, there are also limitations: the models are fundamentally dependent on training datasets being diverse and high-quality; even basic assumptions like "similar

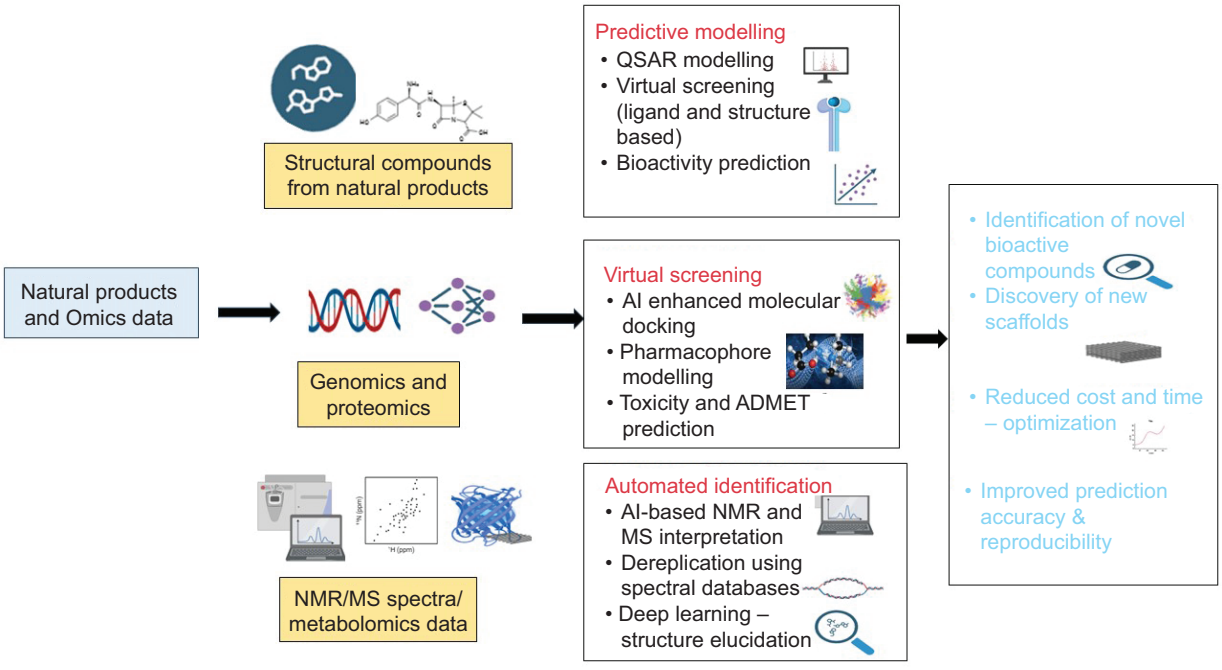


Figure 1. A diagrammatic representation of the integration of AI, ML, and DL approaches in natural product discovery. Predictive modeling allows for bioactivity and toxicity prediction, automated identification of compound accelerates the process of NMR/MS-based dereplication and structure elucidation, AI-assisted virtual screening and docking lead to identification and optimization. Together, these technologies improve efficiency, accuracy, and innovation in natural bioactive discovery. ADMET, absorption, distribution, metabolism, excretion, and toxicity.

structures produce similar properties" are not always valid; and resource-efficient, simplified models may give inaccurate predictions of outcomes due to the high complexity of human biological systems (Vilar et al., 2012; Tiwari et al., 2023; Ocana et al., 2025; Pathan et al., 2025).

3.2. Automated compound identification and structure elucidation

Artificial Intelligence with DL has enabled a new paradigm for compound identification and structural elucidation, by automating and accelerating the interpretation of NMR and MS data. AI has allowed the rapid dereplication of known compounds and identification of unique molecular scaffolds that would otherwise take substantial time for recognition. In turn, these advancements make natural product research and drug discoveries more efficient.

In NMR spectral interpretation, DL methods are being used to streamline many steps of the analytical workflow. For example, pseudo-Siamese convolutional neural networks (pSCNN) can recognize compounds from complex mixtures, even in the presence of overlapping peaks or differences in chemical shifts. The NMR-Solver is an example of integrated software that matches spectra with fragment-based optimization and precalculated experimental and theoretical spectra to provide reliable structure elucidation. Deep learning architectures, such as the JTF-Net, can achieve high-quality spectral reconstruction using undersampled data, and can significantly reduce the time required for data acquisition. This is beneficial, in particular, for metabolomics and protein analysis applications. Furthermore, AI models show better performance than conventional algorithms for peak picking and denoising, and provide clear and reproducible spectra in the presence of noisy or poor-quality data (Luo et al., 2025; Klukowski et al., 2025). Artificial intelligence has transformed dereplication and discovery of new scaffolds in natural product research. By taking complex spectroscopic data from NMR and MS and comparing it to known databases of compounds, AI enables dereplication, saving time and redundancy. In addition, AI can facilitate and direct subsequent targeted enrichment by focusing on fractions that contain new or bioactive compounds. AI models can also assess natural product-likeness to distinguish natural-like structures from synthetic ones. Not only can AI be used for identification, generative AI models can also generate new chemical scaffolds based on natural products, creating new molecular scaffolds to generate drugs with optimized properties (Wei et al., 2022).

Overall, the integration of AI and DL in compound identification and structure elucidation represents a major leap forward in analytical chemistry and drug discovery, enhancing speed, precision, and innovation in exploring nature's chemical diversity.

3.3. High-throughput and virtual screening approaches

The use of AI-accelerated virtual screening transforms drug discovery by employing ML algorithms to quickly screen through thousands of compounds to find the most promising drug targets. High-throughput virtual screening (HTVS) is based on the integration of computational methodologies, such as molecular docking and pharmacophore modeling, to determine the binding of the molecule to a specific biological target. AI can help enhance the accuracy of docking, refine pharmacophore models, and replace the need for in vitro experimental screening. AI-based screening tools apply computational algorithms that predict pharmacokinetic and toxicity properties, lowering the cost of computational analysis and generally improving lead identification. Generative models, such as GANs and VAEs, take a step further by designing an entire New chemical entity based on optimization of the target. In molecular docking, scoring functions trained by AI have directly increased the predictive power of binding affinities, and in further work in pharmacophore modeling, AI even helped to identify salient 3D chemical features essential to establishing biological activity. An integrated-based and pharmacophore-based filtering stage, followed by precise docking-based analyses, provides the highest degree of accuracy in screening hits and optimizing the overall drug development process (Rai et al., 2023; Garg et al., 2024).

4. INTEGRATION OF AI WITH OMICS AND SYSTEMS BIOLOGY

Artificial Intelligence has revolutionized drug discovery and research by leveraging omics and systems biology. Through the combination of data from genomics, proteomics, metabolomics, and epigenomics, AI-powered models can analyze complex biological networks, identify new biomarkers, and predict new therapeutic targets with unprecedented accuracy. ML and DL algorithms can evaluate very large, multidimensional datasets to discover hidden molecular patterns and molecular correlations that are often overlooked by conventional bioinformatic tools. The integration of multi-omics provides deeper understanding of the mechanisms underlying disease, assists with prioritization of drug targets, and speeds up rational drug design (Figure 2).

4.1. Genome mining for biosynthetic gene clusters

Biosynthetic Gene Clusters (BGCs), discovered through genome mining are gene sequences that create secondary metabolites such as antibiotics, antifungals, and anticancer

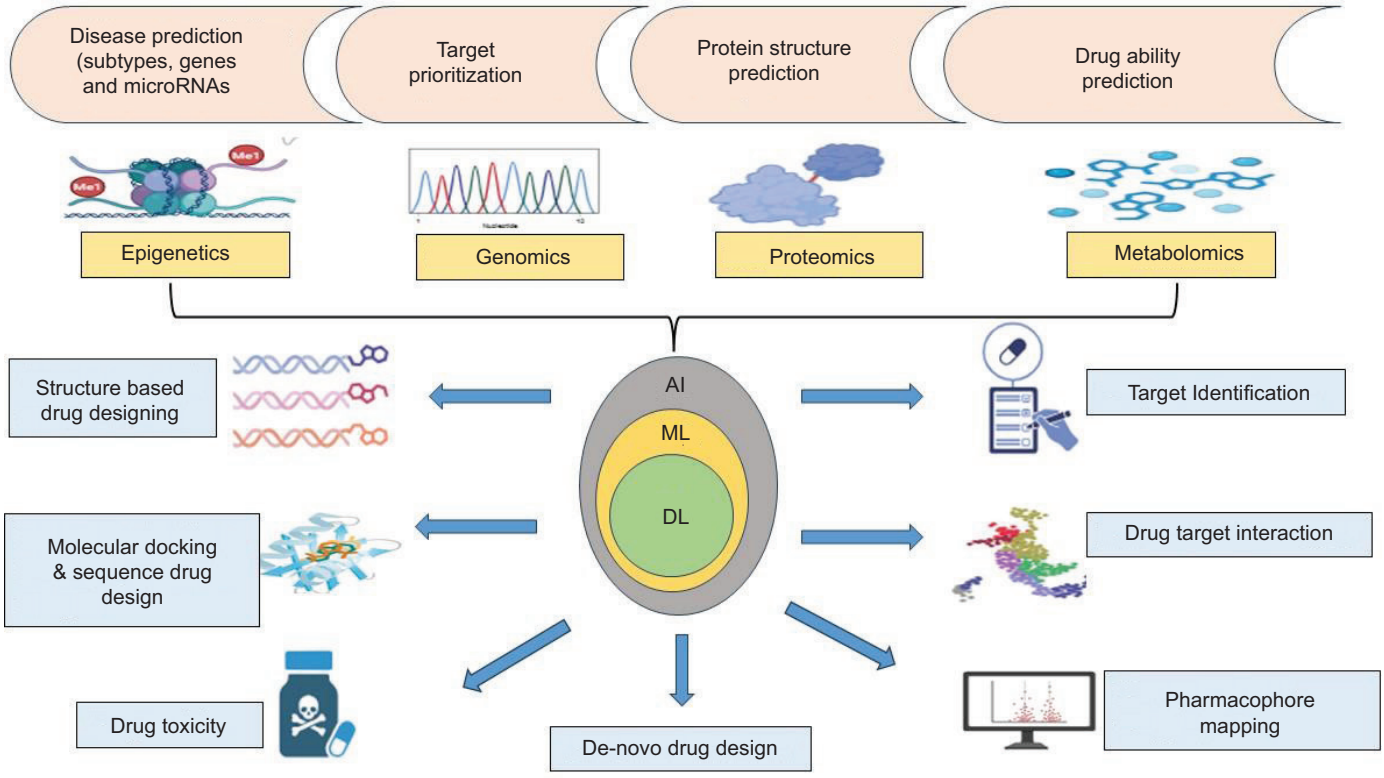


Figure 2. Integration of multi-omics (epigenomics, genomics, proteomics, and metabolomics) with AI, ML, and DL approaches for diverse stages in the discovery of drug, including identification of target, molecular docking, design based on structure and toxicity prediction. AI, Artificial intelligence; DL, deep learning; ML, machine learning.

compounds. The incorporation of ML methods has revolutionized the study of BGCs by identifying new and diverse BGCs beyond the ability of rule-based approaches. Hidden Markov Models (HMMs) identify conserved domains in biosynthetic enzymes (e.g., PKS and NRPS), which are incorporated into software applications like antiSMASH. Deep learning models, specifically CNNs, identify highly complex patterns in the genomic context, leading to improved accuracy and allowing for the discovery of additional BGC types. Traditional ML algorithms, clustered into categories including SVM and Random Forest, identify BGCs based on the presence of genes and domain arrangements. Among the different tools available, antiSMASH uses a combination of rule- and HMM-based approaches to accurately identify and annotate BGCs in a high-throughput manner. DeepBGC employs CNNs to not only identify novel clusters but also predict the bioactivity of the BGC products. Together, these AI-powered tools (PRISM and SMURF) greatly enhance genome mining, along with discoveries of new natural products and other biological components, and provide an understanding of microbial secondary metabolism (Hannigan et al., 2019; Walker et al., 2021; Zhang et al., 2025).

4.2. Metabolomics and cheminformatics integration

Metabolomics and cheminformatics combine biological science and computational sciences to provide a greater understanding about drug mechanism, drug efficacy and drug safety. Metabolomics focuses on the characterization and quantification of metabolites using different techniques, such as LC-MS/MS. Cheminformatics applies a range of computational procedures to evaluate chemical structure by relating molecular properties and can relate them to biological databases. AI-assisted studies may also change how LC-MS/MS data are evaluated due to a range of automated preprocessing steps (i.e., noise removal, peak alignment), metabolite annotation, and prediction of de novo structures, which may assist in bioactive compound and biomarker identification from complex natural extracts more swiftly. Network-based methods, such as molecular networking (GNPS) and NetID, represent the relationships among metabolites based on MS/MS data and thus improve the annotation of unknown metabolites. Network-based approaches, like GNPS and NetID, use MS/MS data to map the relationships between metabolites. This visualization significantly improves the identification

and annotation of previously unknown compounds (Husi et al., 2019; Chen et al., 2021; Xu et al., 2025).

4.3. Multi-Omics data integration and systems pharmacology

The combination multi-omics data and the systems pharmacology applies genomics (genetic profiles), transcriptomics, proteomics, and metabolomics with AI to create map of secondary metabolite biosynthetic pathways for drug discovery. This holistic approach overcomes the limitations of single-omics studies by revealing the complex interactions between different biological layers. The traditional approach for natural product drug discovery can be time-consuming and inefficient. A multi-omics strategy presents a comprehensive and robust way for the discovery of potential therapeutic candidates by combining domain areas of biological information (Table 2).

Systems pharmacology provides an added dimension to the applications of multi-omics, providing a framework to determine the interaction of drug with its targets and the broader biological system. Network-based approaches leverage multi-omics data to generate networks representing gene–gene, protein–protein, and drug–target interactions. This allows researchers to infer drug action across targets or pathways and potentially to predict off-target side effects. AI-assisted pathway mapping is essential for processing the high-dimensional data that arises from multi-omics studies. Datasets that are exceedingly complex can be analyzed by ML and DL algorithms to map biosynthetic pathways, identify correlations, and predict metabolic routes. Existing tools such as graph neural networks (GNNs) and network propagation algorithms can characterize regulatory genes, enzymes, and interactions involved in the biosynthesis of secondary metabolites. This can provide pivotal information in a more rapid manner of a novel and uncharacterized pathways. In a typical workflow for systems pharmacology and multi-omics-driven drug discovery approach begins with data generation, where genomics, transcriptomics, proteomics, and metabolomics data are collected from a natural source under varying conditions. Subsequently data integration takes place in which AI tools unite the various datasets to map correlations across different omics levels. This is followed by pathway mapping, in which integrated networks are visualized and genes, enzymes, and metabolites involved in the biosynthesis of secondary metabolites are inferred. Once potential targets and pathways are identified, target validation is performed experimentally, for example, by using CRISPR-Cas9 to knock out specific genes and assess changes in metabolite production. Finally, these systems-level information determines a potential drug candidate to optimize toward bioactive metabolites for production. Overall, the combination of systems pharmacology with AI-driven multi-omics analysis enables the faster and more precise identification of bioactive compounds with defined biosynthetic pathways, thus leads for the efficient drug discovery Bao et al., 2023; Jiang et al., 2025; Meng et al., 2025).

5. PHARMACOKINETIC AND TOXICOLOGICAL OPTIMIZATION

The pharmaceutical industry has experienced a profound shift paradigmatically due to AI implementations using ML and DL to determine and model properties while reducing the cost and time spent in drug development. Some of the major applications include predicting ADMET properties and pharmacokinetic/pharmacodynamic (PK/PD) models driven by AI, as well as scoring drug-likeness scores based on ML.

Table 2
Overview of different omics approaches and their contributions to drug discovery.

Omics type	Information provided	Contribution to drug discovery	References
Genomics	The organism's complete genetic material including the identification of biosynthetic gene clusters (BGCs).	Allows researchers to identify the potential for an organism to produce certain secondary metabolites, even if they are not expressed under normal lab conditions.	(Palazzotto et al., 2018; Du et al., 2024)
Transcriptomics	The organism's complete set of RNA transcripts at a given time.	Provides insights into the genes being actively expressed, helping to understand the dynamics of secondary metabolite biosynthesis under different environmental conditions or stresses.	(Palazzotto et al., 2018; Du et al., 2024)
Proteomics	The organism complete set of proteins produced.	Identifies and quantifies the enzymes involved in a metabolic pathway. Since many proteins are drug targets, this data is essential to determine the mechanism of action.	(Palazzotto et al., 2018; Du et al., 2024)
Metabolomics	The organism complete set of low-molecular-weight metabolites.	Offers a snapshot of the organism's physiological state and provides direct evidence of the final product of a biosynthetic pathway.	(Palazzotto et al., 2018; Du et al., 2024)

5.1. Predicting ADMET profiles with machine learning

Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) characteristics are essential for evaluation of drug effect and safety. Traditionally, obtaining experimental ADMET data is often a long and expensive process; ML models can provide rapid results, and in silico predictions have been used for the initial filtering of unsuitable drug candidates. Some common ML based molecular property prediction methods include Graph Neural Networks (GNNs), for capturing complex atomic and bonding relationships with the molecule represented as a graph; Deep Neural Networks (DNNs), including multitask models that utilize shared information across multiple ADMET endpoints; Gradient Boosting Algorithms, for example, XGBoost, which accurately predicts the property of compounds; and Ensemble Models, which combine multiple algorithms into one output for improved robustness of prediction. Important predictive platforms such as ADMET-AI, which utilizes GNNs to predict 41 ADMET properties with high accuracy; ADMET Predictor, can forecast over 175 properties in addition to output integration with high-throughput simulation of pharmacokinetic processes; and Deep-PK, applies graph-based DL which needs to predict about 73 endpoints, thus providing accurate and rapid ADMET assessment in the drug discovery (Di Lascio et al., 2023; Siramshetty et al., 2024; Han et al., 2025).

5.2. ML-based drug-likeness scoring of natural compounds

Natural products (NPs) are promising sources of new drugs, but challenges arise in their development due to their structural complexity. ML accounts for NPs and is required for assessing and optimizing natural products based on whether they have suitable physicochemical characteristics for drug development. Standard practices, such as Quantitative Estimation of Drug-likeness (QED), combine simple physicochemical rules but can be too restrictive to identify the most promising candidates. More advanced applications of ML, such as DrugMetric, utilize DL frameworks (namely based on geometries such as variational autoencoders (VAE) and Gaussian mixture models (GMM)) to quantify the chemical space of a compound to approved drugs and outperform the performance of QED for complex natural products. Generative AI models, such as GANs and reinforcement learning strategies, can develop new NP-inspired molecules that optimize for specific properties. Moreover, ML-based classifiers can enhance prioritization in NP discovery through predictions of bioactivity based upon the structural and chemical features of natural compounds with known drugs and their associated protein targets (Jia et al., 2020).

5.3 AI-driven PK/PD modeling

Pharmacokinetic/pharmacodynamic (PK/PD) modeling illustrates the exposure of drug in terms of a biological effect. This field is evolving with advances in AI technology, ultimately providing more effective and efficient approaches relative to prior methods. Hybrid models combining ML with a mechanistic approach, such as physiologically based pharmacokinetic (PBPK) or nonlinear mixed-effect (NLME) models, use parameters predicted by ML-based models to increase precision and minimize experimental effort. Generative AI models, including GANs and VAEs can generate realistic synthetic PK/PD data to augment small datasets that are common with rare diseases or pediatric studies, which enable virtual clinical trials. These approaches can also assist with time-series analysis utilizing neural networks, such as recurrent neural networks (RNNs) and latent-ODE models framework, to capture the complex relationships indicated by irregularly sampled PK/PD data that would have been frequently overlooked by prior modeling approaches. Finally, ML algorithms can facilitate automated model selection, allowing for optimized models and covariate relationship, typically leading to pharmacometrics analysis and identifying more accurately predictive models (Tang et al., 2023; Huang et al., 2024).

6. CASE STUDIES AND SUCCESSFUL APPLICATIONS

AI-assisted discovery has given a notable breakthrough in the discovery novel antibiotics, anticancer drugs, and natural product analogs, driven by DL and generative AI (Table 3).

7. CHALLENGES AND LIMITATIONS

AI-driven drug discovery has fundamentally changed the pharmaceutical industry by greatly improving the identification, design, and optimization of novel therapeutics. Despite these advances, several obstacles limit its applications. A primary issue is the lack of high-quality, diverse datasets that are essential for accurate AI predictions. Drug discovery data are often scarce, fragmented, and biased toward well-studied targets, and the lack of standardization across sources also reduce model generalizability. Natural product datasets face additional hurdles due to the complexity of their molecular structures, inconsistent annotations, and limited experimental activity data, making AI-driven predictions difficult. The "black-box" nature of DL and graph-based models reduces transparency, hindering interpretation, regulatory acceptance and overall trust among scientists. Another issue is the integration of computational predictions with wet lab validation.

Table 3
AI-assisted discovery of novel antibiotics, anticancer agents, and natural product analogs.

Category	Compound/Platform	AI approach	Key outcome/discovery	References
Novel Antibiotics	Halicin	Directed Message Passing Neural Network (D-MPNN)	Discovered structurally novel antibiotic active against multidrug-resistant *Escherichia coli* and *Mycobacterium tuberculosis*	(Bagdad et al., 2024)
	Abaucin	AI screening of 7,500 compounds	Narrow-spectrum antibiotic targeting *Acinetobacter baumannii*, minimal effect on commensals	(Bagdad et al., 2024)
	Two novel antibiotics	Graph Neural Networks (GNNs)	Effective against MRSA & VRE; median MIC 3–4 μg/mL; explainable AI mapped active substructures	(Bagdad et al., 2024)
	β-Lactamase inhibitors	Random Forest & Deep Neural Networks	Non-covalent inhibitors against NDM-1 & CMY-10 β-lactamases; restores β-lactam efficacy	(Bagdad et al., 2024)
Anticancer Agents	Z29077885	AI-integrated screening	STK33 inhibitor; induces apoptosis via STAT3 deactivation	
	EXS-21546	Deep learning optimization	Selective A2A receptor antagonist for lung & renal cancers; Phase I trials	(Albani et al., 2025)
	Ouabain + miR-34	High-throughput AI screening	Revealed synergistic anti-lung cancer effect; 91% model accuracy	(Albani et al., 2025)
	Multimodal ML for pancreatic cancer	Random Forest, DNN, GCN	Screened 1.6 M combinations for pancreatic cancer; identified predictive synergistic pairs	(Albani et al., 2025)
Natural Product Analogues	ReLeaSE Platform	Reinforcement + deep generative neural networks	Designed Janus Kinase 2 inhibitors with desired properties	(Albani et al., 2025)
	RXR modulators	Deep recurrent neural network	Generated novel retinoid X receptor ligands mimicking natural products	(Albani et al., 2025)
	GAN-based synthetic analogues	Generative Adversarial Networks	Created libraries mimicking pharmacologically active scaffolds (alkaloids, polyketides)	(Gangwal et al., 2025)

Many of the AI-designed molecules will be hard to synthesize and assays, test in physiologically relevant system, disconnection between workflows, as well as limitation of the resources can slow down the iterative learning cycles. There are privacy, regulatory restrictions, and ethical concerns with sharing biomedical data. Addressing, these issues require curated and interoperable datasets, standardized data formats, explainable AI frameworks, and closer collaboration between computational researchers and experimental teams. Emerging approaches, such as generative AI integrated with automated synthesis and high-throughput assays, show promise in bridging the in silico–in vitro gap, thus aiding in the faster discovery of safe, effective and novel drugs with less cost and less experimental risk (Merk et al., 2018; Ferreira et al., 2025; Qadri et al., 2025).

8. FUTURE PERSPECTIVES

In the coming era, AI and ML are expected to rapidly accelerate the discovery of natural products-based drugs and therapeutics. Emerging technologies, including Generative AI, Reinforcement Learning, and Quantum ML, are anticipated to transform the design and optimization of new bioactive compounds, enhance drug–target interaction specificity and make it possible to discover new, unexplored chemical scaffolds. Coupling multi-omics data with predictive models driven by AI will provide a systems-level understanding of biosynthetic pathways and enable more effective identification of new natural products. As an added benefit to drug discovery processes, AI-based methodologies could lead to more sustainable and green drug discovery workflows while reducing waste and resource use. As dataset sizes increase and computational models of data become more tractable and interpretable, AI and ML are expected to play an important role in drug development that is more precise, faster, and safer for the environment.

9. CONCLUSION

Artificial Intelligence and ML are being integrated into the field of natural products, marking a fundamental shift in modern drug discovery. These computational tools enable rapid bioactivity prediction, structure elucidation, virtual screening, and genome mining of biosynthetic gene clusters, while overcoming many limitations observed with traditional discovery. Despite persisting obstacles such as lack of data, lack of standardized datasets, model interpretability, and integration with wet-lab validation, AI and ML present

unparalleled means to accelerate the discovery of new natural products and optimize their pharmacokinetic and pharmacodynamics properties. New technologies such as Generative AI, Reinforcement Learning, Quantum ML, and multi-omics integration have the potential to enhance predictive capabilities, facilitate systems-level understanding of natural products biosynthetic systems and facilitate rational drug design approaches. AI-enabled methods also promote sustainable and green drug discovery by decreasing experimental costs, minimizing chemical waste, and promoting a harm-free ecosystem by responsible research practices. As the field of computational approaches continues to progress, the increase in datasets and ethical and regulatory guidance will allow AI and ML to be essential methods in deriving novel, effective, and accurate therapeutics from natural products, ultimately establishing a data-driven and sustainable framework for drug discovery.

ORCID

J Geetha	0009-0004-2941-8424
Nathiya Ranganathan	0009-0000-7760-9306
G Gulothungan	0009-0008-8191-1784
Hitesh Chopra	0000-0001-8867-7603

AUTHOR CONTRIBUTIONS

J Geetha: Research concept and design, Data analysis and interpretation, and Writing of the article; Nathiya Ranganathan: Critical revision and final approval of the article; G Gulothungan: Research concept and design, and writing of the article; Hitesh Chopra: Research concept and design, Writing and final approval of the manuscript.

FUNDING

No funding was received for this manuscript.

CONFLICT OF INTEREST

There is no conflict of interest.

COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any studies involving human participants or animals.

REFERENCES

Albani, F.G., Alghamdi, S.S., Almutairi, M.M., Alqahtani, T., 2025. Artificial intelligence-driven innovations in oncology drug discovery: transforming traditional pipelines and enhancing drug design. Drug Design, Development and Therapy. 19, 5685–5707. https://doi.org/10.2147/DDDT.S509769

Atanasov, A.G., Zotchev, S.B., Dirsch, V.M., Supuran, C.T., 2021. Natural products in drug discovery: advances and opportunities. Nature Reviews Drug discovery. 20(3), 200–216. https://doi.org/10.1038/s41573-020-00114-z

Bagdad, Y., Miteva, M.A., 2024. Recent applications of artificial intelligence in discovery of new antibacterial agents. Advances and Applications in Bioinformatics and Chemistry: AABC. 17, 139–157. https://doi.org/10.2147/AABC.S484321

Bao, H., Zhao, J., Zhao, X., Zhao, C., Lu, X., Xu, G., 2023. Prediction of plant secondary metabolic pathways using deep transfer learning. BMC Bioinformatics. 24(1), 348. https://doi.org/10.1186/s12859-023-05485-9

Blanco-González, A., Cabezón, A., Seco-González, A., Conde-Torres, D., Antelo-Riveiro, P., Piñeiro, Á., Garcia-Fandino, R., 2023. The role of AI in drug discovery: challenges, opportunities, and strategies. Pharmaceuticals (Basel, Switzerland). 16(6), 891. https://doi.org/10.3390/ph16060891

Chen, L., Lu, W., Wang, L., Xing, X., Chen, Z., Teng, X., Zeng, X., Muscarella, A.D., Shen, Y., Cowan, A., McReynolds, M.R., Kennedy, B.J., Lato, A.M., Campagna, S.R., Singh, M., Rabinowitz, J.D., 2021. Metabolite discovery through global annotation of untargeted metabolomics data. Nature Methods. 18(11), 1377–1385. https://doi.org/10.1038/s41592-021-01303-3

Dara, S., Dhamercherla, S., Jadav, S.S., Babu, C.M., Ahsan, M.J., 2022. Machine learning in drug discovery: a review. Artificial Intelligence Review. 55(3), 1947–1999. https://doi.org/10.1007/s10462-021-10058-4

Di Lascio, E., Gerebtzoff, G., Rodriguez-Perez, R., 2023. Systematic evaluation of local and global machine learning models for the prediction of ADME properties. Molecular Pharmaceutics. 20(3), 1758–1767. https://doi.org/10.1021/acs.molpharmaceut.2c00962

Dias, D.A., Urban, S., Roessner, U., 2012. A historical overview of natural products in drug discovery. Metabolites. 2(2), 303–336. https://doi.org/10.3390/metabo2020303

Du, P., Fan, R., Zhang, N., Wu, C., Zhang, Y., 2024. Advances in integrated multi-omics analysis for drug–target identification. Biomolecules. 14(6), 692. https://doi.org/10.3390/biom14060692

Ferreira, F.J., Carneiro, A.S., 2025. AI-driven drug discovery: a comprehensive review. ACS Omega. 10(23), 23889–23903. https://doi.org/10.1021/acsomega.5c00549

Fu, C., Chen, Q., 2025. The future of pharmaceuticals: artificial intelligence in drug discovery and development. Journal of Pharmaceutical Analysis. 15(8), 101248. https://doi.org/10.1016/j.jpha.2025.101248

Gangwal, A., Lavecchia, A., 2025. Artificial intelligence in natural product drug discovery: current applications and future perspectives. Journal of Medicinal Chemistry. 68(4), 3948–3969. https://doi.org/10.1021/acs.jmedchem.4c01257

Garg, P., Singhal, G., Kulkarni, P., Horne, D., Salgia, R., Singhal, S.S., 2024. Artificial Intelligence–driven computational approaches in the development of anticancer drugs. Cancers. 16(22), 3884. https://doi.org/10.3390/cancers16223884

Han, H., Shaker, B., Lee, J. H., Choi, S., Yoon, S., Singh, M., Basith, S., Cui, M., Ahn, S., An, J., Kang, S., Yeom, M.S., Choi, S., 2025. Employing Automated Machine Learning (AutoML) methods to facilitate the in silico ADMET properties prediction. Journal of Chemical Information and Modeling. 65(7), 3215–3225. https://doi.org/10.1021/acs.jcim.4c02122

Hannigan, G.D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., Durcak, J., Wurst, M., Kotowski, J., Chang, D., Wang, R., Piizzi, G., Temesi, G., Hazuda, D.J., Woelk, C.H., Bitton, D.A., 2019. A deep learning genome-mining strategy for biosynthetic gene cluster prediction. Nucleic Acids Research. 47(18), e110. https://doi.org/10.1093/nar/gkz654

Hou, X., Sun, M., Bao, T., Xie, X., Wei, F., Wang, S., 2020. Recent advances in screening active components from natural products based on bioaffinity techniques. Acta Pharmaceutica Sinica B. 10(10), 1800–1813. https://doi.org/10.1016/j.apsb.2020.04.016

https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug, accessed on 22/10/2025.

Huang, Z., Denti, P., Mistry, H., Kloprogge, F., 2024. Machine learning and artificial intelligence in PK-PD modeling: fad, friend, or foe? Clinical Pharmacology and Therapeutics. 115(4), 652. https://doi.org/10.1002/cpt.3165

Husi, H., 2019. Computational biology. Codon Publications, Brisbane (AU).

Jia, C.Y., Li, J. Y., Hao, G.F., Yang, G.F., 2020. A drug-likeness toolbox facilitates ADMET study in drug discovery. Drug Discovery Today. 25(1), 248–258. https://doi.org/10.1016/j.drudis.2019.10.014

Jiang, W., Ye, W., Tan, X., Bao, Y.J., 2025. Network-based multi-omics integrative analysis methods in drug discovery: a systematic review. BioData Mining. 18(1), 27. https://doi.org/10.1186/s13040-025-00442-z

Klukowski, P., Riek, R., Güntert, P., 2025. Machine learning in NMR spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy. 148–149, 101575. https://doi.org/10.1016/j.pnmrs.2025.101575

Luo, Y., Zheng, X., Qiu, M., Gou, Y., Yang, Z., Qu, X., Chen, Z., Lin, Y., 2025. Deep learning and its applications in nuclear magnetic resonance spectroscopy. Progress in Nuclear Magnetic Resonance Spectroscopy. 146–147, 101556. https://doi.org/10.1016/j.pnmrs.2024.101556

Meng, F., Zhang, G., Xiao, W., Mao, Y., Shu, Y., Yang, X., Xu, G., Tang, X., Zhang, M., Liu, Z., Zhang, X., You, S., Wang, B., Yu, Z., Chen, S., Chen, W., 2025. MedMeta: an AI-enabled and genomics-based database for functional profiling of secondary metabolites in medicinal species. Engineering. https://doi.org/10.1016/j.eng.2025.09.007

Merk, D., Grisoni, F., Friedrich, L., Schneider, G., 2018. Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators. Communications Chemistry. 1(1), 68.

Moingeon, P., Kuenemann, M., Guedj, M., 2022. Artificial intelligence-enhanced drug design and development: toward a computational precision medicine. Drug Discovery Today. 27(1), 215–222. https://doi.org/10.1016/j.drudis.2021.09.006

Mungwari, C.P., King'ondu, C.K., Sigauke, P., Obadele, B.A., 2024. Conventional and modern techniques for bioactive compounds recovery from plants. Scientific African. 27(1), e02509. https://doi.org/10.1016/j.sciaf.2024.e02509

Newman, D.J., Cragg, G.M., 2020. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. Journal of Natural Products. 83(3), 770–803.

Obaido, G., Mienye, I.D., Egbelowo, O.F., Emmanuel, I.D., Ogunleye, A., Ogbuokiri, B., Mienye, P., Aruleba, K., 2024. Supervised machine learning in drug discovery and development: algorithms, applications, challenges, and prospects. Machine Learning with Applications. 17, 100576. https://doi.org/10.1016/j.mlwa.2024.100576

Ocana, A., Pandiella, A., Privat, C., Bravo, I., Luengo-Oroz, M., Amir, E., Gyorffy, B., 2025. Integrating artificial intelligence in drug discovery and early drug development: a transformative approach. Biomarker Research. 13(1), 45. https://doi.org/10.1186/s40364-025-00758-2

Palazzotto, E., Weber, T., 2018. Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. Current Opinion in Microbiology. 45, 109–116. https://doi.org/10.1016/j.mib.2018.03.004

Patel, L., Shukla, T., Huang, X., Ussery, D.W., Wang, S., 2020. Machine learning methods in drug discovery. Molecules (Basel, Switzerland). 25(22), 5277. https://doi.org/10.3390/molecules25225277

Pathan, I., Raza, A., Sahu, A., Joshi, M., Sahu, Y., Patil, Y., Raza, M.A., 2025. Revolutionizing pharmacology: AI-powered approaches in molecular modeling and ADMET prediction. Medicine in Drug Discovery. 28, 100223. https://doi.org/10.1016/j.medidd.2025.100223

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., Tekade, R.K., 2021. Artificial intelligence in drug discovery and development. Drug Discovery Today. 26(1), 80–93. https://doi.org/10.1016/j.drudis.2020.10.010

Qadri, Y.A., Shaikh, S., Ahmad, K., Choi, I., Kim, S.W., Vasilakos, A.V., 2025. Explainable artificial intelligence: a perspective on drug discovery. Pharmaceutics. 17(9), 1119. https://doi.org/10.3390/pharmaceutics17091119

Rai, S., Pandey, P., Gupta, A., Sharma, S., 2023. Virtual screening and high-throughput screening in drug discovery. European Journal of Military Studies. 13(4).

Sarkar, C., Das, B., Rawat, V.S., Wahlang, J.B., Nongpiur, A., Tiewsoh, I., Lyngdoh, N.M., Das, D., Bidarolli, M., Sony, H.T., 2023. Artificial intelligence and machine learning technology driven modern drug discovery and development. International Journal of Molecular Sciences. 24(3), 2026. https://doi.org/10.3390/ijms24032026

Sathya, S., Manikandan, S., Rajeswari, V.R., 2025. Natural products with artificial intelligence: a revolution in drug discovery. International Journal of Research in Pharmacy and Allied Science. 4(1), 45–62.

Simoben, C.V., Babiaka, S.B., Moumbock, A.F.A., Namba-Nzanguim, C.T., Eni, D.B., Medina-Franco, J.L., Günther, S., Ntie-Kang, F., Sippl, W., 2023. Challenges in natural product-based drug discovery assisted with *in silico*-based methods. RSC Advances. 13(45), 31578–31594. https://doi.org/10.1039/d3ra06831e

Singh, S., Kumar, R., Payra, S., Singh, S.K., 2023. Artificial intelligence and machine learning in pharmacological research: bridging the gap between data and drug discovery. Cureus. 15(8), e44359. https://doi.org/10.7759/cureus.44359

Siramshetty, V.B., Xu, X., Shah, P., 2024. Artificial intelligence in ADME property prediction. Methods in Molecular Biology (Clifton, N.J.). 2714, 307–327. https://doi.org/10.1007/978-1-0716-3441-7_17

Tang, A., 2023. Machine learning for pharmacokinetic/pharmacodynamic modeling. Journal of Pharmaceutical Sciences. 112(5), 1460–1475. https://doi.org/10.1016/j.xphs.2023.01.010

Thomford, N.E., Senthebane, D.A., Rowe, A., Munro, D., Seele, P., Maroyi, A., Dzobo, K., 2018. Natural products for drug discovery in the 21st century: innovations for novel drug discovery. International Journal of Molecular Sciences. 19(6), 1578. https://doi.org/10.3390/ijms19061578

Tiwari, P.C., Pal, R., Chaudhary, M.J., Nath, R., 2023. Artificial intelligence revolutionizing drug development: exploring opportunities and challenges. Drug Development Research. 84(8), 1652–1663. https://doi.org/10.1002/ddr.22115

Vilar, S., Costanzi, S., 2012. Predicting the biological activities through QSAR analysis and docking-based scoring. Methods in Molecular Biology (Clifton, N.J.). 914, 271–284. https://doi.org/10.1007/978-1-62703-023-6_16

Vora, L.K., Gholap, A.D., Jetha, K., Thakur, R.R.S., Solanki, H.K., Chavda, V.P., 2023. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. 15(7), 1916. https://doi.org/10.3390/pharmaceutics15071916

Walker, A.S., Clardy, J., 2021. A machine learning bioinformatics method to predict biological activity from biosynthetic gene clusters. Journal of Chemical Information and Modeling. 61(6), 2560–2571. https://doi.org/10.1021/acs.jcim.0c01304

Wei, W., Liao, Y., Wang, Y., Wang, S., Du, W., Lu, H., Kong, B., Yang, H., Zhang, Z., 2022. Deep learning-based method for compound identification in NMR spectra of mixtures. Molecules. 27(12), 3653. https://doi.org/10.3390/molecules27123653

Xu, R., Zhu, J., 2025. Unveiling the dark matter of the metabolome: a narrative review of bioinformatics tools for LC-HRMS-Based Compound Annotation. Talanta. 295, 128327. https://doi.org/10.1016/j.talanta.2025.128327

Zhang, L., Liu, Y., Chen, K., Yue, Q., Wang, C., Xie, L., Molnár, I., Xu, Y., 2025. Genome mining of tailoring enzymes from biosynthetic gene clusters for synthetic biology: a case study with fungal methyltransferases. Metabolic Engineering. 92, 125–135. https://doi.org/10.1016/j.ymben.2025.08.001