Medical : Jurnal Kesehatan Dan Kedokteran
Volume 3, Issue 2, July-December 2026
e-ISSN: 3064-6472 (online)
https://jurnal.sinesia.id/index.php/Medical-JKK/index

Emerging Role of Artificial Intelligence in Natural Product-Based Drug Discovery

Yelfi Anwar[1], Syifa Zafira Nasution[2], Lelly Farida[3], Andrianopsyah Mas Jaya Putra[4]
yelfi.anwar@uta45jakarta.ac.id
[1,3]Faculty of Pharmacy, Universitas 17 Agustus 1945 Jakarta, Indonesia
[2]Universitas Pancasila, Indonesia
[3]Research Center for Process Technology National Research and Innovation Agency (BRIN), Indonesia

Info Article
|Submitted: 4 September 2026 **|Revised**: 2 October 2026 **|Accepted**: 4 November 2026
|Published: 4 November 2026
How to cite: Yelfi Anwar, etc., "Emerging Role of Artificial Intelligence in Natural Product-Based Drug Discovery", *Medical : Jurnal Kesehatan Dan Kedokteran*, Vol. 3, Issue 2, July-December, 2026, P. 376-394.

ABSTRACT

Background: Natural products remain an important source of therapeutic agents, but their discovery is constrained by chemical complexity, time-consuming isolation, difficult structure elucidation, and high attrition rates. Artificial intelligence (AI), including machine learning, deep learning, graph neural networks, and generative AI, offers computational approaches to address these challenges. Method: This narrative review examines peer-reviewed, English-language literature published between 2021 and 2026, retrieved from PubMed, Scopus, ScienceDirect, and Google Scholar. The selected literature was analyzed using thematic qualitative synthesis to identify AI applications, reported contributions, and persistent limitations across the natural product discovery pipeline. Problem Identification: Despite growing interest in AI-assisted drug discovery, the extent to which current methods address the limitations of conventional approaches remains uncertain. Key concerns include data quality, model reliability, interpretability, and the need for experimental validation. Results: AI applications cover target identification, biosynthetic gene cluster mining, virtual screening, structure elucidation, network pharmacology, generative molecular design, and absorption, distribution, metabolism, excretion, and toxicity (ADMET) prediction. These methods can help prioritize candidate compounds and predict potential biological activities. However, data scarcity and bias, inconsistent evaluation practices, limited interpretability, and insufficient experimental validation remain significant barriers. Explainable AI, multi-omics integration, and self-driving laboratories represent promising developments, although their practical contributions require further validation. Conclusion: AI can support and accelerate selected stages of natural product-based drug discovery but cannot replace experimental testing or the expertise of natural product chemists and pharmacologists.

Keywords: *artificial intelligence; natural products; drug discovery; machine learning*

Introduction

Natural products have historically occupied a pivotal role in drug discovery, with approximately 50% of approved drugs derived directly or indirectly from natural sources (Mullowney et al., 2023). Their structural diversity and evolutionary biological relevance make them uniquely valuable for therapeutic development compared to purely synthetic compound libraries. Critically, natural products explore regions of chemical space inaccessible to synthetic combinatorial chemistry, possessing stereocenters, ring systems, and functional group patterns that have

This work is licensed under CC BY-SA 4.0

been refined by millions of years of evolutionary pressure to interact with biological macromolecules (Yuan et al., 2025).

Despite this rich pharmacological legacy, traditional approaches to natural product-based drug discovery are notoriously labor-intensive, time-consuming, and limited by the complexity of natural product mixtures. The isolation and structural elucidation of a single compound can require months to years of work, the overall attrition rate from hit-to-drug remains extremely high, and issues of bioavailability and scalability of synthesis frequently impede clinical translation (Abbas et al., 2024; Gangwal & Lavecchia, 2025).

Recent advances in artificial intelligence (AI) have introduced transformative opportunities across pharmaceutical research. AI techniques-particularly machine learning (ML), deep learning (DL), graph neural networks (GNNs), and large language models (LLMs)-enable the integrated analysis of large, heterogeneous datasets and complex biological systems, facilitating faster, more efficient, and more cost-effective drug discovery processes (Ferreira & Carneiro, 2025; Zhang et al., 2025). The convergence of AI with genomics, metabolomics, proteomics, and structural biology creates unprecedented opportunities to identify, characterize, and optimize novel bioactive compounds before ever stepping into a laboratory.

Based on the above background, the central problem addressed in this review can be stated as follows: how do current AI methodologies overcome the limitations of conventional methods at each stage of the natural product drug discovery pipeline, from target identification through lead optimization, and what challenges must still be resolved for widespread adoption. This review therefore explores the emerging role of AI in natural product-based drug discovery, focusing on recent innovations, applications, challenges, and future perspectives as of 2022-2026. Special attention is given to how AI methodologies are reshaping each stage of the drug discovery pipeline, from genomic prediction of undiscovered natural scaffolds through to preclinical lead optimization. The intended benefit of this review is to provide researchers, pharmacy practitioners, and health professionals with a comprehensive map of how AI is reshaping natural product research and where future efforts should be directed.

Research Methods

This study employed a narrative literature review to examine the role of artificial intelligence (AI) in natural product-based drug discovery. Literature published between 2021 and 2026 was considered, with searches conducted through PubMed, Scopus, ScienceDirect, and Google Scholar. The search strategy used combinations of the keywords “artificial intelligence,” “machine learning,” “deep learning,” “natural products,” and “drug discovery.” Publications were selected

based on their relevance to AI methodologies, applications, challenges, and emerging developments in natural product research.

The inclusion criteria comprised English-language, peer-reviewed journal articles addressing AI applications in one or more stages of the natural product drug discovery pipeline. Both original research articles and review articles were considered. The scope included target identification, biosynthetic gene cluster mining, virtual screening, bioactivity prediction, structure elucidation, network pharmacology, generative AI, absorption, distribution, metabolism, excretion, and toxicity (ADMET) prediction, lead optimization, explainable AI (XAI), and self-driving laboratories. Articles lacking substantive discussion of AI or computational approaches to drug discovery, non-peer-reviewed publications, and conference proceedings without sufficient supporting evidence were excluded.

A total of 465 records were identified through database searches and additional citation searching. After removing 95 duplicate records, 370 articles remained for screening based on their titles and abstracts. Of these, 280 articles were excluded, leaving 90 articles for full-text eligibility assessment. Following a detailed review, 65 articles were excluded because they did not specifically address AI applications in natural product-based drug discovery, consisted only of abstracts, short reviews, or editorials, or lacked accessible full text or did not meet the English-language requirement. Ultimately, 25 articles met the inclusion criteria and were included in the qualitative analysis. These articles were examined to identify AI methodologies, applications across the natural product drug discovery pipeline, implementation challenges, and future research opportunities.

The selected literature was organized into seven thematic areas: (1) natural products as a foundation for drug discovery; (2) AI methodologies; (3) AI applications across the drug discovery pipeline; (4) current challenges; (5) explainable AI; (6) self-driving laboratories and autonomous research; and (7) future perspectives. The synthesis drew on the literature cited throughout the Results and Discussion section, including studies by Mullowney et al. (2023), Abbas et al. (2024), Tang et al. (2024), Ekins (2024), Ferreira and Carneiro (2025), Yuan et al. (2025), Zhang et al. (2025), Gangwal and Lavecchia (2025), Ang and Choo (2026), and other relevant sources listed in the manuscript's bibliography.

The analysis was descriptive and qualitative. For each thematic area, the literature was examined in terms of the AI methods employed, their reported applications, the evidence supporting their performance, their limitations, and their potential implications for natural product discovery. Particular attention was given to data availability, model interpretability, experimental validation, reproducibility, and the distinction between established applications and emerging technologies. Findings were synthesized narratively rather than statistically, and no meta-

analysis was performed. This review was based on previously published literature and did not involve primary data collection or laboratory experimentation.

Results and Discussion

The literature review indicates that the application of artificial intelligence in natural product-based drug discovery has expanded rapidly since 2022, spanning nearly every stage of the drug discovery pipeline. The results of the thematic synthesis are presented in the following subsections, beginning with the position of natural products as a foundation for drug discovery, an overview of AI methodologies, their application across the discovery pipeline, current challenges, and future perspectives for the field.

2.1 Natural Products as a Foundation for Drug Discovery

Natural products remain a cornerstone of drug discovery owing to their unparalleled structural complexity, evolutionary optimization for biological interactions, and often multi-target pharmacological activity (Mullowney et al., 2023). The structural diversity of natural products far exceeds that of synthetic compound libraries, providing unique chemical scaffolds that interact with a diverse range of biological targets. Notable examples include paclitaxel (Taxol) from the Pacific yew tree, artemisinin from Artemisia annua, and the statin drug lovastatin from Aspergillus terreus, all of which emerged from natural product-based programs and have saved or improved millions of lives (Yuan et al., 2025).

Key advantages of natural products in drug discovery include their structural diversity, representing billions of years of evolutionary refinement; their inherent biological compatibility with living systems; their capacity for multi-target activity, interacting simultaneously with multiple molecular targets; and their access to unexplored regions of chemical novelty that synthetic compound libraries cannot replicate (Abbas et al., 2024; Mullowney et al., 2023).

Advancements in omics technologies-including genomics, transcriptomics, proteomics, and metabolomics-have substantially expanded access to previously unexplored natural compounds, enabling the discovery of novel bioactive molecules at a scale impossible to conceive of in earlier decades (Yuan et al., 2025). Despite these advantages, significant barriers to drug development persist: low extraction yields, complex isolation procedures, poor bioavailability, structural elucidation difficulties, and a pervasive "rediscovery problem" where known compounds are repeatedly isolated from new sources instead of truly novel scaffolds (Ang & Choo, 2026; Gangwal & Lavecchia, 2025).

2.2 Overview of Artificial Intelligence Methodologies

Artificial intelligence encompasses a broad spectrum of computational methodologies that enable systems to learn from data and make predictions or generate novel outputs (Bellini et al., 2022). In natural product drug discovery, four

primary categories of AI approaches have demonstrated significant promise: (1) supervised learning, used for Quantitative Structure-Activity Relationship (QSAR) modeling in bioactivity prediction, toxicity classification, ADMET profiling, and target binding affinity estimation; (2) unsupervised learning, used for clustering of compound libraries, dimensionality reduction for chemical space visualization, and dereplication of known compounds from metabolomic data; (3) deep learning, comprising Convolutional Neural Networks (CNNs) for spectral analysis, Recurrent Neural Networks (RNNs) for sequence-based prediction, and Transformer architectures for molecular representation learning; and (4) generative AI, comprising Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), diffusion models, and reinforcement learning frameworks for de novo molecular design.

These AI methods enable prediction of biological activity with high accuracy, toxicity assessment before experimental testing, comprehensive drug-target interaction analysis, identification of structure-activity relationships, and increasingly, the autonomous design of entirely novel molecular scaffolds (Tang et al., 2024; Zhang et al., 2025).

Table 1. Representative AI Models and Their Applications in Natural Product Drug Discovery

AI Category	Representative Models/Algorithms	Primary Function	Example Applications	References
Machine Learning (ML)	Random Forest (RF), Support Vector Machine (SVM), XGBoost	Pattern recognition and predictive modeling	QSAR modeling, toxicity prediction, ADMET profiling	Abbas et al., 2024; Ferreira & Carneiro, 2025
Deep Learning (DL)	CNNs, RNNs, Transformers	Automated feature extraction from complex datasets	Spectral interpretation, bioactivity prediction, molecular representation learning	Mullowney et al., 2023; Yuan et al., 2025
Graph Neural Networks (GNNs)	Message Passing Neural Networks (MPNNs), Graph	Molecular graph representation and structure-	Drug-target interaction prediction, molecular	Tang et al., 2024; Zhang et al., 2025

AI Category	Representative Models/Algorithms	Primary Function	Example Applications	References
	Convolutional Networks (GCNs)	property analysis	property prediction	
Generative AI	GANs, VAEs, Diffusion Models	Generation of novel molecular structures	De novo drug design, scaffold optimization	Ferreira & Carneiro, 2025; Tang et al., 2024
Large Language Models (LLMs)	ChemBERTa, MolBERT, BioGPT, Mol-GPT	Chemical language understanding and molecular sequence prediction	SMILES generation, retrosynthesis planning, multitarget drug design	Ferreira & Carneiro, 2025; Zhang et al., 2025
Explainable AI (XAI)	SHAP, LIME, attention-based models	Model interpretability and mechanistic insight extraction	Identification of bioactive molecular motifs and decision transparency	Gangwal & Lavecchia, 2025; Lavecchia, 2025

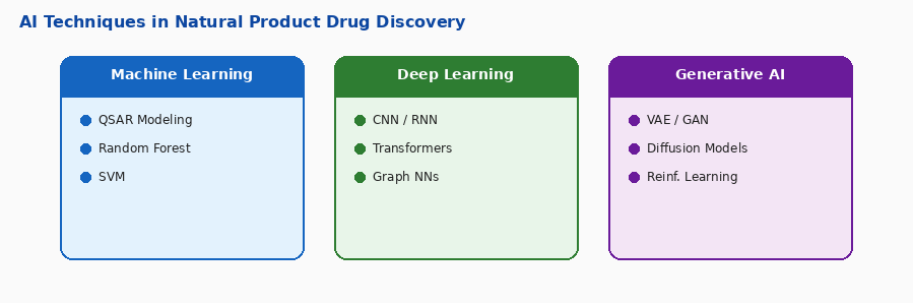


Figure 1. Overview of major AI technique categories applied in natural product drug discovery, spanning machine learning, deep learning, and generative AI approaches.

A particularly notable methodological advance is the development of Graph Neural Networks (GNNs), which represent molecules as graphs with atoms as nodes and bonds as edges, capturing topological molecular information in a mathematically natural way that fingerprint-based methods cannot achieve (Tang et al., 2024).

The field has also witnessed the rapid adaptation of large language models (LLMs) to chemical applications. Models such as MolBERT, ChemBERTa, and Mol-GPT encode molecular SMILES strings or graph representations, enabling property

prediction, structure generation, and retrosynthesis planning with unprecedented accuracy (Ferreira & Carneiro, 2025). Diffusion models-a class of generative AI originally developed for image synthesis-have shown particular promise in generating three-dimensional molecular conformations for improved docking accuracy (Zhang et al., 2025).

2.3 AI Applications Across the Drug Discovery Pipeline

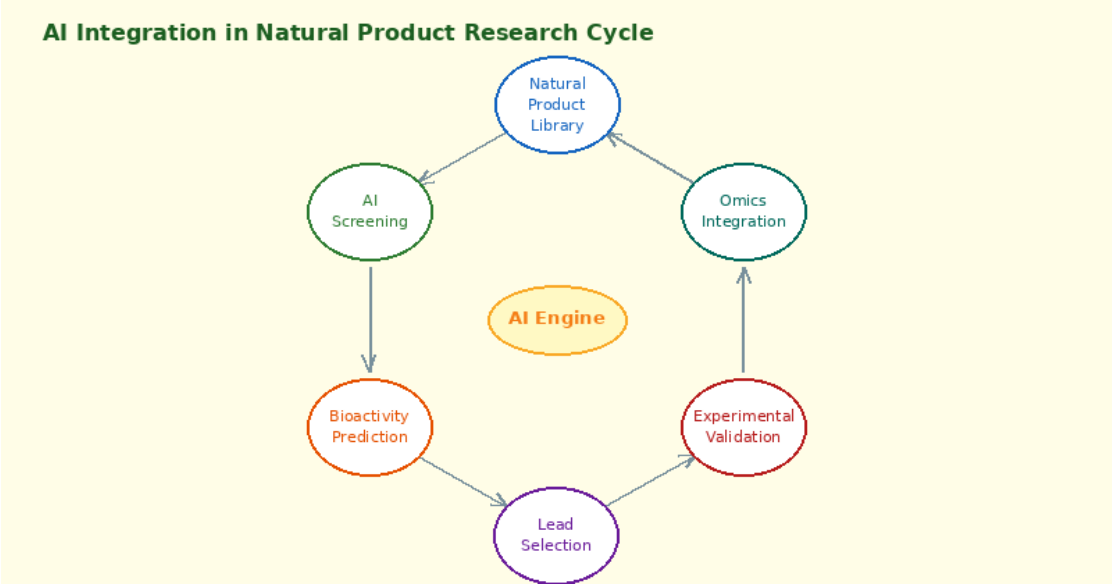


Figure 2. The AI-integrated natural product research cycle, illustrating the iterative loop from compound library screening through omics integration, enabling continuous discovery refinement.

2.3.1 Target Identification and Validation

Target identification-determining which biological molecules are causally involved in a disease-is the critical first step in drug discovery. AI-powered bioinformatics tools, particularly those leveraging network pharmacology and multi-omics data integration, have dramatically enhanced this process for natural product research (Abbas et al., 2024). Machine learning models trained on large-scale gene expression, protein-protein interaction, and pathway databases can identify the most likely targets of complex natural product mixtures, even when the individual bioactive components are yet unknown.

AI tools analyze biosynthetic gene clusters (BGCs) in microbial and plant genomes to predict novel natural products before laboratory isolation (Ang & Choo, 2026; Mullowney et al., 2023). Platforms such as antiSMASH (version 8.0), DeepBGC, and PRISM leverage machine learning-including hidden Markov models and deep neural networks-to identify and classify BGCs, predict the chemical structures of encoded metabolites, and guide metabolic engineering strategies (Ang & Choo, 2026). Genome mining with antiSMASH has revealed that only approximately 10% of BGCs in Streptomyces species are expressed under standard

laboratory culture conditions, suggesting a vast untapped reservoir of bioactive chemistry (Ang & Choo, 2026).

2.3.2 Virtual Screening and Activity Prediction

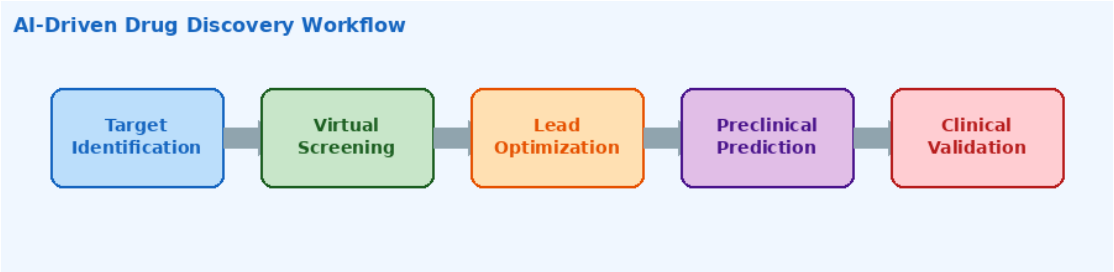


Figure 3. Sequential AI-driven drug discovery workflow from target identification through clinical validation, illustrating the iterative nature of each stage.

AI models represent a paradigm shift in the screening of natural product libraries. Machine learning classifiers and deep neural networks can evaluate millions of compounds in silico within hours, compared to weeks or months required for traditional high-throughput screening (Ferreira & Carneiro, 2025; Gangwal & Lavecchia, 2025). These models predict key pharmacological properties including anticancer activity against specific cancer cell lines, antimicrobial properties against resistant pathogens, anti-inflammatory effects, and complete ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profiles (Abbas et al., 2024).

Several AI-predicted natural compounds have been subsequently validated experimentally, demonstrating real-world applicability. A notable example includes the AI-assisted identification of novel antimicrobial peptides from plant sources with potent activity against drug-resistant bacteria (Mullowney et al., 2023). Additionally, QSAR-assisted virtual screening has enabled the prioritization of lead-like molecules from marine and microbial natural sources for antitumor and antibiotic drug discovery (Fontanella et al., 2026). Recent deep learning implementations achieve area-under-curve (AUC) values exceeding 0.90 in distinguishing active from inactive natural compounds across multiple therapeutic target classes (Ferreira & Carneiro, 2025).

2.3.3 AI-Assisted Structure Elucidation

One of the most time-consuming steps in natural product research is structure elucidation. AI-assisted spectral interpretation tools have dramatically accelerated this process by applying deep learning to NMR, mass spectrometry (MS), infrared (IR), and UV spectral data (Mullowney et al., 2023; Yuan et al., 2025). Transformer-based models trained on large spectral databases can now perform automated NMR spectral assignment in minutes, reducing a task that previously required an expert chemist days or weeks.

Mass spectral dereplication-the process of identifying already-known compounds to prioritize genuinely novel compounds for further investigation-has

been transformed by AI-powered tools integrated with metabolomics databases such as GNPS, MassBank, and COCONUT. These systems rapidly match experimental fragmentation patterns against curated libraries while flagging structurally novel compounds that do not match known entries, fundamentally changing research prioritization workflows (Ang & Choo, 2026; Yuan et al., 2025).

Table 2. Major Public Databases and Computational Platforms Used in AI-Driven Natural Product Drug Discovery

Database/Platform	Primary Data Type	Main Function	Relevance to AI-Based Drug Discovery	References
COCONUT	Natural product structures	Curated open-access natural product database	Training datasets for ML and generative AI models	Fontanella et al., 2026; Yuan et al., 2025
LOTUS	Linked natural product information	Integration of structure, taxonomy, and references	Knowledge graph construction and data mining	Ang & Choo, 2026; Yuan et al., 2025
ChEMBL	Bioactivity and pharmacological data	Drug-target interaction database	QSAR modeling and activity prediction	Abbas et al., 2024; Ferreira & Carneiro, 2025
GNPS	Mass spectrometry data	Molecular networking and metabolomics analysis	AI-assisted dereplication and spectral annotation	Mullowney et al., 2023; Yuan et al., 2025
MassBank	Mass spectral records	Spectral comparison and compound identification	Deep learning-based spectral prediction	Ang & Choo, 2026; Mullowney et al., 2023

Database/Platform	Primary Data Type	Main Function	Relevance to AI-Based Drug Discovery	References
antiSMASH	Biosynthetic gene clusters (BGCs)	Genome mining for secondary metabolites	Prediction of novel natural products from genomic data	Ang & Choo, 2026; Mullowney et al., 2023
DeepBGC	Genomic sequence data	AI-driven BGC prediction and classification	Automated identification of biosynthetic pathways	Ang & Choo, 2026
PRISM	Secondary metabolite gene clusters	Structural prediction of encoded metabolites	Computational discovery of cryptic natural products	Ang & Choo, 2026; Mullowney et al., 2023

2.3.4 Network Pharmacology and Multi-Target Analysis

Understanding how natural compounds interact with biological targets is fundamental to drug development, particularly given the polypharmacological nature of most phytochemicals. AI-powered network pharmacology tools integrate multi-omics data, protein-protein interaction networks, and knowledge graphs to map compound-target relationships comprehensively (Yuan et al., 2025). This approach is particularly important for complex herbal medicines, where therapeutic effects frequently arise from synergistic interactions among multiple bioactive components acting on multiple disease-related targets simultaneously.

Knowledge graph-based AI systems can now integrate compound structures, known biological activities, gene ontologies, disease phenotypes, and clinical trial outcomes into unified frameworks, enabling hypothesis-driven identification of novel therapeutic applications for known natural product scaffolds (Abbas et al., 2024). These approaches have yielded promising leads in oncology, neurodegeneration, and antimicrobial resistance research.

2.3.5 Generative AI and De Novo Drug Design

Generative AI models have opened entirely new avenues for drug design by creating novel chemical structures inspired by or derived from natural product scaffolds (Tang et al., 2024; Zhang et al., 2025). Several primary generative

architectures are currently employed in this context: Generative Adversarial Networks (GANs) generate realistic molecular structures with desired properties by training a generator-discriminator pair in adversarial fashion; Variational Autoencoders (VAEs) enable continuous latent-space exploration of chemical diversity, enabling gradient-based optimization of molecular properties; reinforcement learning frameworks optimize molecular properties through iterative reward-based generation; and graph neural networks model molecular structure as graphs for property-directed generation (Tang et al., 2024).

Large language model-based approaches such as MTMol-GPT have further extended capabilities to multi-target drug design, generating compounds predicted to bind multiple disease-relevant targets simultaneously-a property particularly relevant for natural product-inspired scaffolds (Zhang et al., 2025). Diffusion model-based three-dimensional structure generation has emerged as a powerful complement to sequence-based methods, enabling more accurate binding pose prediction and improved docking-guided optimization (Ferreira & Carneiro, 2025).

2.3.6 ADMET Prediction and Lead Optimization

Lead optimization-the iterative improvement of a promising candidate compound to enhance its drug-like properties while retaining biological activity-has been fundamentally transformed by AI. Deep learning models trained on large pharmacokinetic datasets now predict ADMET properties with accuracy approaching experimental measurements, enabling virtual optimization cycles that would previously require extensive laboratory work (Abbas et al., 2024; Gangwal & Lavecchia, 2025).

Integration of AI-predicted ADMET profiles into generative design pipelines has enabled the development of compound libraries simultaneously optimized for potency, selectivity, metabolic stability, and low toxicity. This multi-objective optimization, once intractable with traditional computational approaches, is now routinely accomplished through reinforcement learning frameworks that reward compounds satisfying multiple pharmacological criteria simultaneously (Ferreira & Carneiro, 2025; Zhang et al., 2025).

2.4 Current Challenges

Despite its transformative potential, the application of AI in natural product drug discovery faces several significant challenges that must be addressed to realize its full promise (Abbas et al., 2024; Ferreira & Carneiro, 2025; Gangwal & Lavecchia, 2025), as summarized in Table 3.

Table 3. Current Challenges and Proposed Solutions in AI-Driven Natural Product Drug Discovery

Challenge	Description	Proposed Solution	References
Data Availability	Natural product databases are far smaller and less consistently annotated than synthetic chemistry databases; ChEMBL exceeds 2 million compounds while the largest curated natural product databases contain orders of magnitude fewer entries	Federated learning; data augmentation; open collaborative databases (COCONUT, LOTUS, UNPD)	Fontanella et al., 2026; Yuan et al., 2025
Model Interpretability	Black-box deep learning models are difficult for chemists and biologists to interpret, limiting acceptance and the ability to extract mechanistic insights from predictions	Explainable AI (XAI); SHAP values; attention-based visualizations; surrogate models	Lavecchia, 2025
Experimental Validation	AI predictions require expensive and time-consuming in vitro and in vivo experimental validation, creating a bottleneck in translating computational findings to actionable drug candidates	High-throughput screening; organ-on-chip platforms; self-driving laboratories	Ekins, 2024; Ferreira & Carneiro, 2025
Data Heterogeneity	Natural product research generates highly heterogeneous data: spectral records, bioassay results, genomic sequences, and metabolomic profiles use incompatible formats and measurement standards	Standardized ontologies; multi-omics integration frameworks; knowledge graph alignment	Ang & Choo, 2026; Yuan et al., 2025

Challenge	Description	Proposed Solution	References
Reproducibility	Lack of standardized benchmarking datasets and evaluation protocols makes it difficult to compare AI model performance across research groups and leads to inconsistent reporting	Standardized benchmark datasets; open-source model sharing; reproducibility checklists	Sourlos et al., 2024; Zhang et al., 2025
Chemical Space Bias	AI models trained predominantly on well-characterized compound classes (e.g., polyketides, terpenoids) may perform poorly when applied to structurally unusual or underexplored natural product families	Transfer learning; few-shot learning; domain adaptation strategies	Santana et al., 2021; Tang et al., 2024

Critical Insight: The "data bottleneck" remains the primary limiting factor in AI-driven natural product research. Unlike synthetic chemistry databases such as ChEMBL (exceeding 2 million curated bioactivity records), even the most comprehensive natural product databases such as COCONUT or UNPD contain an order of magnitude fewer consistently annotated entries. Addressing this gap through collaborative, open-access data sharing initiatives, standardized reporting frameworks, and automated extraction of natural product data from literature using NLP tools represents perhaps the single highest-impact opportunity to accelerate the field.

2.5 Explainable AI (XAI) in Natural Product Drug Discovery

The "black-box" nature of deep learning models-particularly their inability to readily explain why a specific prediction was made-is recognized as a major barrier to trust and adoption in pharmaceutical research (Lavecchia, 2025). Explainable AI (XAI) has emerged as a crucial solution for enhancing transparency, reliability, and mechanistic insight in AI-driven drug discovery pipelines.

Key XAI methodologies now applied in natural product research include SHapley Additive exPlanations (SHAP), which uses cooperative game theory to quantify the contribution of individual molecular features to model predictions; Local Interpretable Model-Agnostic Explanations (LIME), which builds locally accurate surrogate models; attention-based visualization, which highlights which molecular substructures or spectral regions drive model outputs; saliency maps;

and concept whitening methods that align model representations with human-understandable chemical concepts (Lavecchia, 2025).

Regulatory considerations further drive the demand for XAI. The US FDA's 2025 draft guidance and the European Medicines Agency's reflection paper both underscore the need for transparent, traceable, and context-aware validation of AI systems used in drug discovery (Lavecchia, 2025). XAI approaches that provide mechanistic rationales alongside predictions-not merely numerical confidence scores-are increasingly required to satisfy regulatory expectations and build the scientific community's trust in AI-generated hypotheses.

For natural product research specifically, XAI enables researchers to identify which structural motifs in complex phytochemical scaffolds are responsible for predicted bioactivities, guiding rational optimization and semi-synthetic modification strategies. This transforms AI from a black-box oracle into a mechanistic hypothesis-generation tool that actively advances chemical understanding (Gangwal & Lavecchia, 2025; Lavecchia, 2025).

2.6 Self-Driving Laboratories and Autonomous Research

The emerging paradigm of self-driving laboratories (SDLs)-autonomous systems combining AI agents with robotic platforms for closed-loop experimental design, execution, and analysis-holds particular transformative promise for natural product drug discovery (Ekins, 2024). These systems can autonomously formulate hypotheses, design experiments, interpret results, and iteratively refine their research strategies, effectively compressing discovery timelines from years to months in some applications.

Contemporary SDL implementations such as LUMI-lab (integrated molecular modeling and robotics), MicroCycle, and AI Science Factory platforms demonstrate that fully automated design-make-test-learn cycles are achievable for pharmaceutical compound libraries (Ekins, 2024). In a landmark demonstration, the LUMI-lab platform screened over 1,700 lipid nanoparticle formulations and discovered a novel class of brominated lipid enhancers of mRNA delivery-a finding that would have been prohibitively time-consuming using traditional screening approaches.

For natural product research, SDL approaches are particularly well-suited to the iterative optimization of extraction conditions, bioassay-guided fractionation, and semi-synthetic modification of natural scaffolds. Integration of AI-driven hypothesis generation with robotic compound handling and automated analytical detection creates a fundamentally new mode of natural product research capable of operating at scales and speeds inaccessible to human researchers (Ekins, 2024).

2.7 Future Perspectives

The trajectory of AI in natural product drug discovery points toward an increasingly integrated, automated, and intelligent research ecosystem (Gangwal &

Lavecchia, 2025; Mullowney et al., 2023). Several key future directions are emerging as priority areas for the field: multi-omics integration, combining genomics, transcriptomics, proteomics, and metabolomics data within unified AI frameworks to create holistic models of natural compound bioactivity and biosynthesis, enabling prediction of yet-undiscovered compound classes from genomic data alone (Ang & Choo, 2026; Yuan et al., 2025); explainable AI (XAI) expansion, through the development of interpretable models that provide mechanistic insights alongside predictions (Lavecchia, 2025); personalized phytotherapy, the AI-driven identification of natural product combinations tailored to individual patient genetic profiles, microbiome compositions, and disease characteristics (Gangwal & Lavecchia, 2025; Yuan et al., 2025); green chemistry AI, integrating AI with sustainable extraction, biosynthetic pathway engineering, and synthetic biology approaches to maximize yield while minimizing environmental impact (Ang & Choo, 2026; Mullowney et al., 2023); self-driving laboratories, fully automated robotic platforms guided by AI agents in closed-loop workflows (Ekins, 2024); and foundation models for chemistry, large pre-trained molecular foundation models trained on all available chemical, spectroscopic, bioactivity, and genomic data, serving as universal platforms for downstream natural product discovery tasks (Ferreira & Carneiro, 2025; Zhang et al., 2025).

Recent industrial collaborations between pharmaceutical companies and AI firms highlight the growing commercial importance of these approaches. It is estimated that AI-assisted drug discovery could reduce the average cost of bringing a new drug to market by up to 40%, while simultaneously increasing the probability of clinical success (Chakraborty et al., 2024; Ferreira & Carneiro, 2025). The first wave of fully AI-designed drug candidates entering clinical trials-including Insilico Medicine's ISM001-055 showing positive Phase IIa results in idiopathic pulmonary fibrosis-demonstrates that AI-driven drug discovery has progressed from computational demonstration to clinical reality (Ekins, 2024).

2.8. Critical Assessment of Evidence and Translational Readiness

The literature indicates that the contribution of artificial intelligence (AI) to natural product-based drug discovery varies across applications and levels of experimental validation. Machine learning for compound prioritization, computational dereplication, and biosynthetic gene cluster analysis provides practical support for identifying promising candidates and reducing unnecessary experimental work. However, the reliability of these approaches depends on the quality of chemical and biological annotations, the diversity of training datasets, and the availability of independent validation. Natural product research presents additional challenges because extracts may contain multiple constituents, while the identity and abundance of individual bioactive compounds may remain uncertain (Saldívar-González et al., 2022; Mullowney et al., 2023).

A distinction should also be made between computational performance and biological or clinical validation. Virtual screening, target prediction, network pharmacology, and ADMET prediction can help rank candidates and generate testable hypotheses, but their predictions do not independently establish biological activity, mechanism of action, safety, or therapeutic efficacy. Generative AI similarly offers opportunities to explore chemical structures inspired by natural products, although the novelty, synthetic feasibility, biological activity, and safety of generated molecules require experimental assessment. These limitations are consistent with broader evaluations of AI-assisted drug discovery, which emphasize the need for reliable reference data, reproducibility, and human expertise throughout the development pipeline (Hasselgren & Oprea, 2024).

Emerging technologies, including multimodal foundation models, explainable AI, and self-driving laboratories, may strengthen the integration of computational prediction and experimental research. Nevertheless, these approaches should be presented as developing capabilities rather than universally established solutions. Their value must be evaluated using transparent performance measures, independent datasets, and experimental confirmation. Evidence of improved computational prioritization should not automatically be interpreted as evidence of reduced development costs, shorter clinical timelines, or higher drug approval rates. A recent review of AI applications in natural product discovery likewise identifies experimental validation, dataset limitations, and reproducibility as continuing challenges to translation (Lavecchia, 2026).

Taken together, the findings support a complementary role for AI in natural product-based drug discovery. Its strongest current contribution is to assist researchers in prioritizing compounds, integrating complex datasets, and formulating hypotheses for subsequent investigation. Progress toward therapeutic applications will depend on connecting these computational outputs with chemical characterization, biological assays, pharmacological evaluation, and, where justified, clinical studies. This distinction provides a more balanced interpretation of AI's present capabilities while identifying the evidence needed to establish its long-term impact.

Conclusion

Artificial intelligence is fundamentally transforming natural product-based drug discovery by enabling faster, more efficient, and data-driven approaches to identifying novel therapeutics. The integration of machine learning, deep learning, graph neural networks, and generative AI models with traditional natural product research has demonstrated remarkable success across virtual screening, structure elucidation, biosynthetic gene cluster mining, target identification, network pharmacology, and de novo drug design, directly addressing the problem identified

in the Introduction regarding how AI overcomes the limitations of conventional methods.

While challenges related to data availability and heterogeneity, model interpretability, experimental validation, and reproducibility remain significant, the field is advancing at an unprecedented pace. The development of explainable AI frameworks, expansion of curated natural product databases, integration with multi-omics technologies, and the emergence of self-driving laboratories will further enhance the impact of AI in pharmaceutical research.

Critically, the future of natural product drug discovery is not one where AI replaces the expertise of natural product chemists and pharmacognosists, but rather one where AI serves as a powerful collaborative tool that amplifies human creativity, accelerates experimental cycles, and illuminates previously inaccessible regions of chemical and biological space. The convergence of centuries of accumulated natural product knowledge with twenty-first-century AI capabilities represents one of the most promising frontiers in pharmaceutical science.

Recommendations

Based on this review, researchers and academic institutions are encouraged to prioritize the establishment of openly accessible, standardized, and curated national natural product databases, in order to reduce the data-scarcity bottleneck that continues to constrain the development of AI models in this field.

Model developers and bioinformatics practitioners are encouraged to make explainable AI (XAI) a standard component of every AI pipeline for natural product drug discovery, so that predictions can be mechanistically interpreted and more readily accepted by chemists, pharmacologists, and regulators.

Regulators and policymakers are encouraged to continue developing clear guidance on the validation and use of AI systems in drug development, in line with the frameworks already being initiated by international regulatory bodies.

Future researchers are encouraged to conduct broader experimental validation studies of AI-predicted natural product compounds, including through collaboration with self-driving laboratories, so that computational findings can be more rapidly translated into preclinically and clinically tested drug candidates.

References

Abbas, M. K. G., Rassam, A., Karamshahi, F., Abunora, R., & Abouseada, M. "The Role of AI in Drug Discovery." ChemBioChem 25, no. 14 (2024): e202300816. https://doi.org/10.1002/cbic.202300816.

Ang, M. Y., & Choo, S. W. "The Computational Revolution in Natural Product Research: A Data-Driven Roadmap for Next-Generation Drug Development." Biology 15, no. 8 (2026): 632. https://doi.org/10.3390/BIOLOGY15080632.

Artificial intelligence for natural product drug discovery and development: Current landscape, applications, and future directions. (2025). Intelligence-Based Medicine, 12, Article 100316. https://doi.org/10.1016/j.ibmed.2025.100316

Bellini, V., Cascella, M., Cutugno, F., Russo, M., Lanza, R., Compagnone, C., & Bignami, E. "Understanding Basic Principles of Artificial Intelligence: A Practical Guide for Intensivists." Acta Biomedica 93, no. 5 (2022): e2022297. https://doi.org/10.23750/abm.v93i5.13626.

Chakraborty, C., Bhattacharya, M., Lee, S. S., Wen, Z. H., & Lo, Y. H. "The Changing Scenario of Drug Discovery Using AI to Deep Learning: Recent Advancement, Success Stories, Collaborations, and Challenges." Molecular Therapy Nucleic Acids 35, no. 3 (2024): 102295. https://doi.org/10.1016/j.omtn.2024.102295.

Ekins, S. "The Lab of the Future: Self-Driving Labs for Molecule Discovery." GEN Biotechnology 3, no. 2 (2024): 83-86.

Ferreira, F. J. N., & Carneiro, A. S. "AI-Driven Drug Discovery: A Comprehensive Review." ACS Omega 10, no. 23 (2025): 23889-23903.

Fontanella, F., D'Alessandro, T., Nardone, E., De Stefano, C., Vicidomini, C., & Roviello, G. N. "Artificial Intelligence for Natural Products Drug Discovery in Neurodegenerative Therapies: A Review." Biomolecules 16, no. 1 (2026): 129. https://doi.org/10.3390/biom16010129.

Gangwal, A., & Lavecchia, A. "Artificial Intelligence in Natural Product Drug Discovery: Current Applications and Future Perspectives." Journal of Medicinal Chemistry 68, no. 4 (2025): 3948-3969.

Hasselgren, C., & Oprea, T. I. (2024). Artificial intelligence for drug discovery: Are we there yet? Annual Review of Pharmacology and Toxicology, 64, 527–550. https://doi.org/10.1146/annurev-pharmtox-040323-040828

Lavecchia, A. "Explainable Artificial Intelligence in Drug Discovery: Bridging Predictive Power and Mechanistic Insight." Wiley Interdisciplinary Reviews: Computational Molecular Science 15, no. 5 (2025): e70049. https://doi.org/10.1002/wcms.70049.

Lavecchia, A. (2026). Artificial intelligence-driven natural product drug discovery: From computational genome mining to clinical translation. Medicinal Research Reviews. https://doi.org/10.1002/med.70098

Mullowney, M. W., Duncan, K. R., Elsayed, S. S., Garg, N., van der Hooft, J. J. J., Martin, N. I., Meijer, D., et al. "Artificial Intelligence for Natural Product Drug Discovery." Nature Reviews Drug Discovery 22, no. 11 (2023): 895-916. https://doi.org/10.1038/s41573-023-00774-7.

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), Article 1119. https://doi.org/10.3390/pharmaceutics17091119

Saldívar-González, F. I., Aldas-Bulos, V. D., Medina-Franco, J. L., & Plisson, F. (2022). Natural product drug discovery in the artificial intelligence era. Chemical Science, 13(6), 1526–1546. https://doi.org/10.1039/D1SC04471K

Santana, K., do Nascimento, L. D., Lima e Lima, A., Damasceno, V., Nahum, C., Braga, R. C., & Lameira, J. "Applications of Virtual Screening in Bioprospecting: Facts, Shifts, and Perspectives to Explore the Chemo-Structural Diversity of Natural Products." Frontiers in Chemistry 9 (2021): 662688. https://doi.org/10.3389/fchem.2021.662688.

Sourlos, N., Vliegenthart, R., Santinha, J., Klontzas, M. E., Cuocolo, R., Huisman, M., & van Ooijen, P. "Recommendations for the Creation of Benchmark Datasets for Reproducible Artificial Intelligence in Radiology." Insights into Imaging 15, no. 1 (2024): 248. https://doi.org/10.1186/s13244-024-01833-2.

Tang, X., Dai, H., Knight, E., Wu, F., Li, Y., Li, T., & Gerstein, M. "A Survey of Generative AI for De Novo Drug Design: New Frontiers in Molecule and Protein Generation." Briefings in Bioinformatics 25, no. 4 (2024): bbae338. https://doi.org/10.1093/bib/bbae338.

Yuan, X., Yang, X., Pan, Q., Luo, C., Luan, X., & Zhang, H. "Artificial Intelligence in Natural Products Research." Chinese Journal of Natural Medicines 23, no. 11 (2025): 1342-1357. https://doi.org/10.1016/S1875-5364(25)60902-2.

Zhang, O., Lin, H., Zhang, X., Wang, X., Wu, Z., Ye, Q., Zhao, W., et al. "Graph Neural Networks in Modern AI-Aided Drug Discovery." Chemical Reviews 125, no. 20 (2025): 10001-10103. https://doi.org/10.1021/acs.chemrev.5c00461.