IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131122

Artificial Intelligence in Pharmacognosy: Transforming Natural Product Drug Discovery

Ms.SrushtI Maske[1], Ms Laxmi Maske[2], Ms Sayali Nanaware[3], Prof.Dr Nitin Waghmode[4]

Dattakala college of pharmacy swami chincholi Tal Daund Dis Pune Maharashtra India[1,2,3,4]

Abstract: Natural products contain chemical diversity, but their discovery is impeded by complex mixtures, repeated discoveries, vague and uncertain therapeutic targets, and validation. Artificial intelligence (AI) can link metabolomics, genomics, and other molecular data and streamline the pharmacognosy workflow. This narrative review summarizes the papers published between 2020 to 2025 on chemical classification, dereplication, genome mining, molecular property prediction, docking, and the assessment of absorption, distribution, metabolism, excretion and toxicity (ADMET). AI, along with deep learning and other emerging generative models, can predict structures, annotate compounds, and classify them based on their expected activities. Reassuringly, some structures have been annotated with high confidence, but deep learning models remain untested. Thus, in the case of AI, main reliance should be placed on experimental validation.

Keywords: artificial intelligence; pharmacognosy; natural products; machine learning; drug discovery; phytochemistry; deep learning.

1. INTRODUCTION

1.Computational Methods and Pharmacognosy

The importance of naturally-occurring products as starting materials for pharmaceuticals is recognized. Nevertheless, isolation, identification and evaluation of bioactivities of constituents from various source materials remain challenging. Conventional methods of extraction and fractionation tend to reconstruct previously described metabolites. The value of computational approaches depend on credible and authenticated source materials and reliable databases, and must be independently verified. [1-11]

2. ARTIFICIAL INTELLIGENCE IN MODERN PHARMACOGNOSY

Data for pharmacognosy, integrating molecular descriptors and biomass, is synthesized with genomes and bioassay data. Through machine learning, representations and generation of possible scenarios for metabolites and scaffolds are accomplished. High-quality and standardized data and metadata of naturally-occurring products must be assured. Predictions must define their confidence intervals and their applicability domain, and provide proofs and traceability of models. [12-22]

3. AI-ASSISTED NATURAL PRODUCT IDENTIFICATION AND SCREENING

The number of features derived from untargeted LC-MS/MS data is usually higher than the number of features that can be isolated in a lab. Methods based on machine learning, spectral similarity and molecular networking abound to aid metabolite prioritization and dereplication. However, in many cases, verifying the gene products and biosynthesis require further experiments. A screening workflow requires a thorough evaluation integrating structure and function.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131122

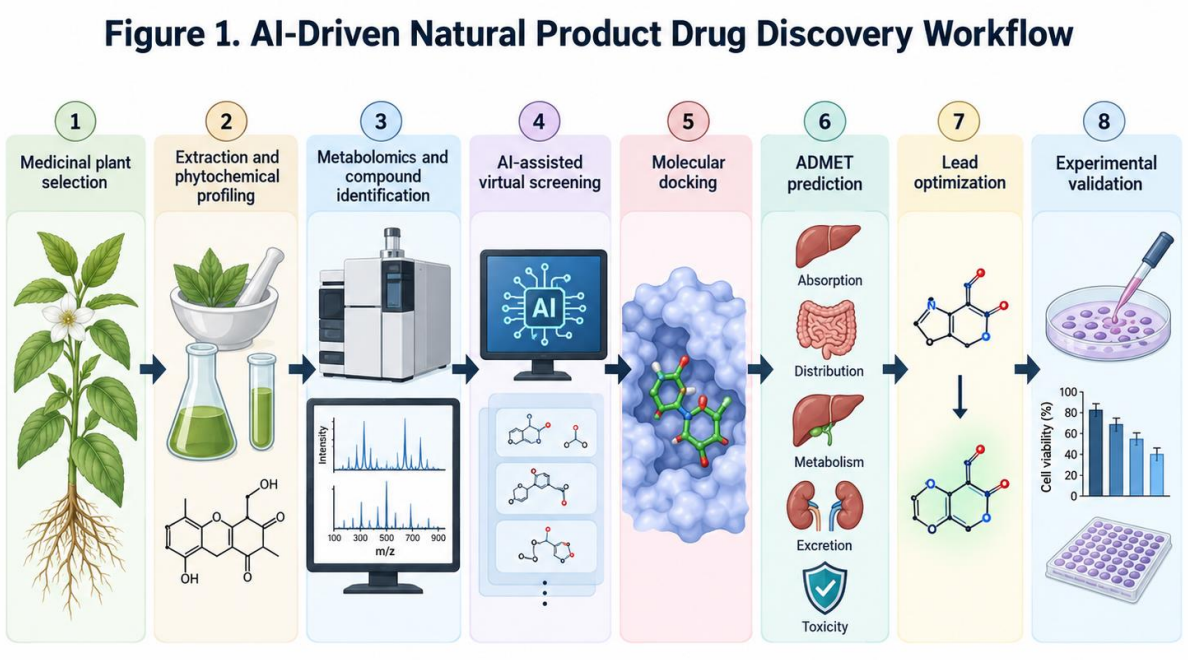


Figure 1. AI-driven natural-product discovery workflow: plant selection; extraction and phytochemical profiling; metabolomics; AI screening; docking; ADMET; lead optimization; experimental validation.

4. MACHINE LEARNING AND DEEP LEARNING IN DRUG DISCOVERY

QSAR models, graph neural networks, and language models all predict molecular properties and activities. However, these models often generalize poorly to unusal natural product structures or stereoisomers. Performance can be overstated by random data splits in the presence of structural analogs; scaffold separation and/or external testing is preferred. Analogues generated by AI require a safety review and lab evaluation for chemical feasibility. [35-47]

5. AI-BASED MOLECULAR DOCKING AND DRUG-TARGET INTERACTION ANALYSIS

Neural networks can prioritize drug targets, and in some cases, predict binding affinities. However, docking scores do not reflect experimental binding constants. Many factors including protein flexibility and conformation, protonation, solvation, and stereochemistry can influence predicted interactions. Large glycosylated species can pose problems. In the presence of unfavorable interactions, orthogonal biochemical assays, counter-screening, and appropriate simulations are needed before target engagement can be asserted. [48-58]

6. AI IN PHARMACOKINETICS, TOXICITY PREDICTION, AND LEAD OPTIMIZATION

ADMET models can predict challenges related to solubility and permeability, as well as toxicity. In lead optimization, multiple objectives including potency, exposure, safety and other considerations must be balanced. Models trained on synthetic compounds are unlikely to perform well for natural products. Calibration, uncertainty reporting, consideration of metabolites, and confirmatory experiments are essential. [59-70]

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 11, November 2026

DOI: 10.17148/IARJSET.2026.131122

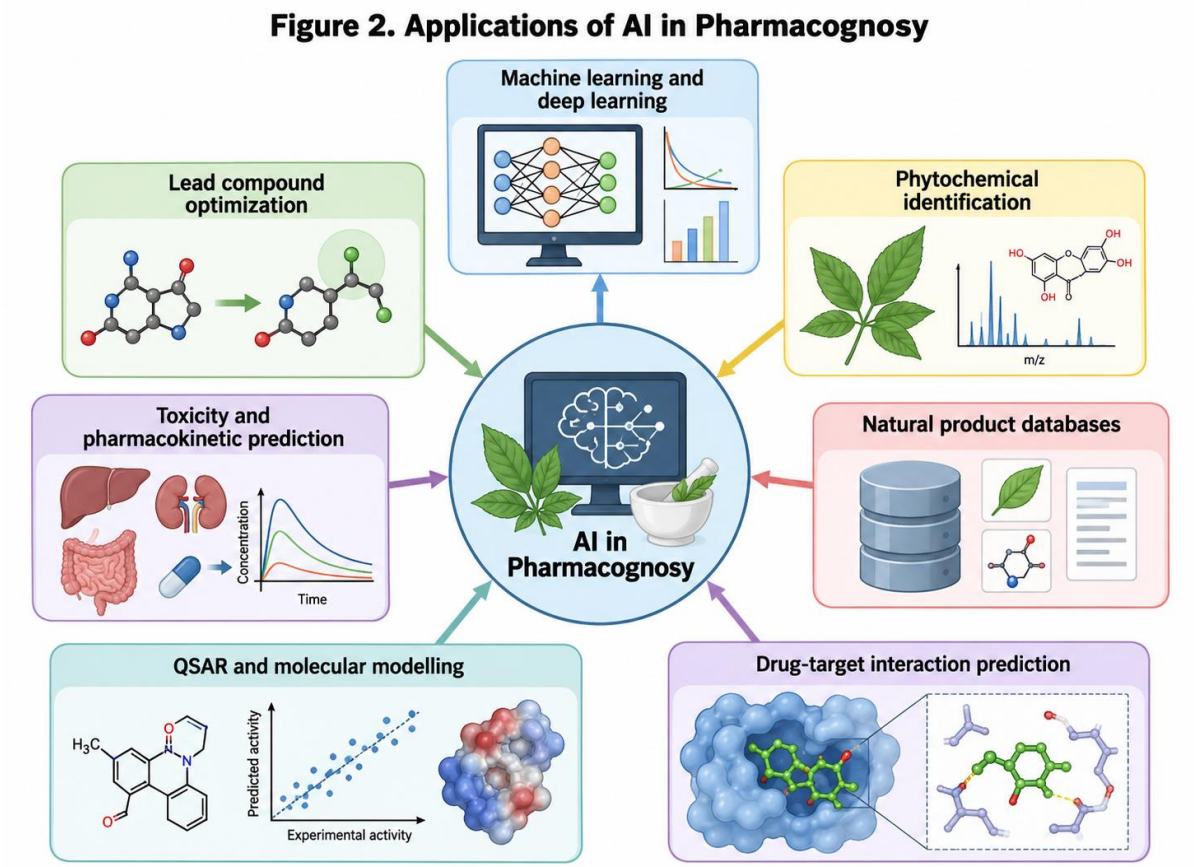


Figure 2. AI applications in pharmacognosy: learning algorithms, natural-product databases, phytochemical identification, target prediction, QSAR modelling, toxicity evaluation and lead prioritization.

7. COMPARATIVE ANALYSIS OF CONVENTIONAL AND AI-DRIVEN PHARMACOGNOSY

Traditional assays have historically produced measured biological outcomes and indicated chemical identities. However, these techniques suffer from low throughput. AI based technologies have significantly improved throughput of ranking and compound association and are limited by sampling bias and incorrect labeling. Novel discoveries require proposed comparisons to independently validated hits, novel chemicals and assessments of assay burden, costs and benefits.

8. CURRENT CHALLENGES AND FUTURE PROSPECTS

Natural products and rare taxa are poorly represented in current natural product and chemo-taxonomic databases. Challenges include poor negative results, promiscuity and heterogeneity of biological samples, poor assay standardization and uncertainty in structure-activity relationships. Transfer learning is one approach to address these challenges. Prioritization of scaffold-based partitions, along with prospective validation and estimating uncertainty should be standard practices in model development. The sustainable use of natural resources along with authentication and expert evaluation of experimental design remain essential. A model should not be rewarded for simply reproducing a previously known scaffold. True innovation and the more important benchmarks should be inspired by biological selectivity and demonstrated through experimentation. Good practice dictates that reports of similar or related models should disclose whether duplicates were removed, compounds were standardized, the source organism was defined, 3D structures of compounds of interest were included and the assays employed were compared and/or validated. [81-97]

9. CONCLUSION

AI has the potential to integrate diverse types of pharmacognosy evidence, and provide insight into relevant candidates. However, meaningful outcomes will be the result of a representative dataset, and independent validation of the predictions. AI will complement and enhance traditional analytical chemistry and medicinal chemistry practices. [98-100]

DOI: 10.17148/IARJSET.2026.131122

REFERENCES

[1]. Newman et al. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod. 2020. https://doi.org/10.1021/acs.jnatprod.9b01285
[2]. Atanasov et al. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov. 2021. https://doi.org/10.1038/s41573-020-00114-z
[3]. Mullowney et al. Artificial intelligence for natural product drug discovery. Nat Rev Drug Discov. 2023. https://doi.org/10.1038/s41573-023-00774-7
[4]. Chigozie et al. Application of artificial intelligence in bioprospecting for natural products for biopharmaceutical purposes. BMC Artif Intell. 2025. https://doi.org/10.1186/s44398-025-00004-7
[5]. Sorokina et al. COCONUT online: Collection of Open Natural Products database. J Cheminform. 2021. https://doi.org/10.1186/s13321-020-00478-9
[6]. Capecchi et al. Classifying natural products from plants, fungi or bacteria using COCONUT and machine learning. J Cheminform. 2021. https://doi.org/10.1186/s13321-021-00559-3
[7]. Kim et al. NPClassifier: a deep neural network-based structural classification tool for natural products. J Nat Prod. 2021. https://doi.org/10.1021/acs.jnatprod.1c00399
[8]. Stravs et al. MSNovelist: de novo structure generation from mass spectra. Nat Methods. 2022. https://doi.org/10.1038/s41592-022-01486-3
[9]. Blin et al. antiSMASH 7.0: improved predictions for detection, regulation and chemical structures. Nucleic Acids Res. 2023. https://doi.org/10.1093/nar/gkad344
[10]. Terlouw et al. MIBiG 3.0: a community-driven effort to annotate biosynthetic gene clusters. Nucleic Acids Res. 2023. https://doi.org/10.1093/nar/gkac1049
[11]. Chandrasekhar et al. COCONUT 2.0: comprehensive overhaul of the natural products database. Nucleic Acids Res. 2024. https://doi.org/10.1093/nar/gkae1063
[12]. Bajorath et al. Deep machine learning for computer-aided drug design. Front Drug Discov. 2022. https://doi.org/10.3389/fddsv.2022.829043
[13]. Gangwal et al. Generative artificial intelligence in drug discovery: framework and recent advances. Front Pharmacol. 2024. https://doi.org/10.3389/fphar.2024.1331062
[14]. Hasselgren et al. Artificial intelligence for drug discovery: are we there yet?. Annu Rev Pharmacol Toxicol. 2024. https://doi.org/10.1146/annurev-pharmtox-040323-040828
[15]. Singh et al. Application of artificial intelligence in drug design: a review. Comput Biol Med. 2024. https://doi.org/10.1016/j.compbiomed.2024.108810
[16]. Visan et al. Integrating artificial intelligence for drug discovery and drug delivery. Life. 2024. https://doi.org/10.3390/life14020233
[17]. Bender et al. Artificial intelligence in drug discovery: realistic impacts and illusions. Drug Discov Today. 2021. https://doi.org/10.1016/j.drudis.2020.12.009
[18]. Arnold et al. Applications of machine learning in microbial natural product drug discovery. Expert Opin Drug Discov. 2023. https://doi.org/10.1080/17460441.2023.2251400
[19]. Stokes et al. A deep learning approach to antibiotic discovery. Cell. 2020. https://doi.org/10.1016/j.cell.2020.01.021
[20]. Periwal et al. Bioactivity assessment of natural compounds using target similarity machine learning. PLoS Comput Biol. 2022. https://doi.org/10.1371/journal.pcbi.1010029
[21]. David et al. Molecular representations in AI-driven drug discovery: a review and practical guide. J Cheminform. 2020. https://doi.org/10.1186/s13321-020-00460-5
[22]. Chen et al. Artificial intelligence for drug discovery: resources, methods and applications. Mol Ther Nucleic Acids. 2023. https://doi.org/10.1016/j.omtn.2023.02.019
[23]. Blaschke et al. REINVENT 2.0: an AI tool for de novo drug design. J Chem Inf Model. 2020. https://doi.org/10.1021/acs.jcim.0c00915
[24]. Bagal et al. MolGPT: molecular generation using a transformer-decoder model. J Chem Inf Model. 2021. https://doi.org/10.1021/acs.jcim.1c00600
[25]. Bai et al. Deep learning methods for de novo drug design and molecular dynamics. WIREs Comput Mol Sci. 2022. https://doi.org/10.1002/wcms.1581
[26]. Bai et al. MolAICal: 3D drug design by artificial intelligence and classical algorithms. Brief Bioinform. 2021. https://doi.org/10.1093/bib/bbaa161
[27]. Baillif et al. Deep generative models for 3D molecular structure. Curr Opin Struct Biol. 2023. https://doi.org/10.1016/j.sbi.2023.102566
[28]. Kotsias et al. Direct steering of de novo molecular generation with conditional recurrent networks. Nat Mach Intell. 2020. https://doi.org/10.1038/s42256-020-0174-5

DOI: 10.17148/IARJSET.2026.131122

[29]. Krenn et al. SELFIES and the future of molecular string representations. Patterns. 2022. https://doi.org/10.1016/j.patter.2022.100588
[30]. Krenn et al. Self-referencing embedded strings: a robust molecular string representation. Mach Learn Sci Technol. 2020. https://doi.org/10.1088/2632-2153/aba947
[31]. Karpov et al. Transformer-CNN: a Swiss knife for QSAR modeling and interpretation. J Cheminform. 2020. https://doi.org/10.1186/s13321-020-00423-w
[32]. Jiang et al. Graph neural networks for molecular representation in drug discovery. J Cheminform. 2021. https://doi.org/10.1186/s13321-020-00479-8
[33]. Xiong et al. Graph neural networks for automated de novo drug design. Drug Discov Today. 2021. https://doi.org/10.1016/j.drudis.2021.02.011
[34]. Grisoni et al. Bidirectional molecule generation with recurrent neural networks. J Chem Inf Model. 2020. https://doi.org/10.1021/acs.jcim.9b00943
[35]. Grisoni et al. Combining generative artificial intelligence and on-chip synthesis for drug design. Sci Adv. 2021. https://doi.org/10.1126/sciadv.abg3338
[36]. Flam-Shepherd et al. Language models can learn complex molecular distributions. Nat Commun. 2022. https://doi.org/10.1038/s41467-022-30839-x
[37]. Moret et al. Chemical language models leverage molecular structure and bioactivity. Nat Commun. 2023. https://doi.org/10.1038/s41467-022-35692-6
[38]. Mouchlis et al. Advances in de novo drug design: conventional to machine learning methods. Int J Mol Sci. 2021. https://doi.org/10.3390/ijms22041676
[39]. Gao et al. Synthesizability of molecules proposed by generative models. J Chem Inf Model. 2020. https://doi.org/10.1021/acs.jcim.0c00174
[40]. Goldman et al. Defining levels of automated chemical design. J Med Chem. 2022. https://doi.org/10.1021/acs.jmedchem.2c00334
[41]. Graff et al. Accelerating high-throughput virtual screening through active learning. Chem Sci. 2021. https://doi.org/10.1039/d0sc06805e
[42]. Huang et al. DeepPurpose: deep learning library for drug-target interaction prediction. Bioinformatics. 2020. https://doi.org/10.1093/bioinformatics/btaa1005
[43]. Nguyen et al. GraphDTA: predicting drug-target binding affinity with graph neural networks. Bioinformatics. 2021. https://doi.org/10.1093/bioinformatics/btaa921
[44]. Hassan-Harrirou et al. RosENet: improving binding affinity prediction using 3D neural networks. J Chem Inf Model. 2020. https://doi.org/10.1021/acs.jcim.0c00075
[45]. Kwon et al. AK-score: protein-ligand binding affinity with 3D convolutional networks. Int J Mol Sci. 2020. https://doi.org/10.3390/ijms21228424
[46]. Rifaioglu et al. DEEPScreen: drug-target prediction using convolutional neural networks. Chem Sci. 2020. https://doi.org/10.1039/c9sc03414e
[47]. Monteiro et al. DTITR: drug-target binding affinity prediction with transformers. Comput Biol Med. 2022. https://doi.org/10.1016/j.compbiomed.2022.105772
[48]. Dehghan et al. TripletMultiDTI: multimodal drug-target interaction prediction. Expert Syst Appl. 2023. https://doi.org/10.1016/j.eswa.2023.120754
[49]. Palhamkhani et al. DeepCompoundNet: multimodal compound-protein interaction prediction. J Biomol Struct Dyn. 2023. https://doi.org/10.1080/07391102.2023.2291829
[50]. Özçelik et al. Structure-based drug discovery with deep learning. ChemBioChem. 2022. https://doi.org/10.1002/cbic.202200776
[51]. Kwapien et al. Additivity and nonadditivity for machine learning models in drug design. ACS Omega. 2022. https://doi.org/10.1021/acsomega.2c02738
[52]. Skinnider et al. Chemical language models navigate sparsely populated chemical space. Nat Mach Intell. 2021. https://doi.org/10.1038/s42256-021-00368-1
[53]. Méndez-Lucio et al. De novo generation of hit-like molecules from gene expression signatures. Nat Commun. 2020. https://doi.org/10.1038/s41467-019-13807-w
[54]. Moret et al. Beam search for automated design of novel ROR ligands. Angew Chem Int Ed. 2021. https://doi.org/10.1002/anie.202104405
[55]. Lim et al. Scaffold-based molecular design with a graph generative model. Chem Sci. 2020. https://doi.org/10.1039/c9sc04503a
[56]. Lin et al. Generative adversarial networks in drug design and discovery. Molecules. 2020. https://doi.org/10.3390/molecules25143250
[57]. Polykovskiy et al. Molecular sets: a benchmark for molecular generation models. Front Pharmacol. 2020. https://doi.org/10.3389/fphar.2020.565644

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IARJSET.2026.131122

[58]. van Deursen et al. GEN: highly efficient SMILES explorer using autodidactic networks. J Cheminform. 2020. https://doi.org/10.1186/s13321-020-00425-8
[59]. van Tilborg et al. Exposing molecular machine learning limitations with activity cliffs. J Chem Inf Model. 2022. https://doi.org/10.1021/acs.jcim.2c01073
[60]. Sun et al. Graph convolutional networks for computational drug development. Brief Bioinform. 2020. https://doi.org/10.1093/bib/bbz042
[61]. Sunseri et al. Libmolgrid: GPU-accelerated molecular gridding for deep learning. J Chem Inf Model. 2020. https://doi.org/10.1021/acs.jcim.9b01145
[62]. Liu et al. AI-based language models powering drug discovery and development. Drug Discov Today. 2021. https://doi.org/10.1016/j.drudis.2021.06.009
[63]. Martinelli et al. Generative machine learning for de novo drug discovery: systematic review. Comput Biol Med. 2022. https://doi.org/10.1016/j.compbiomed.2022.105403
[64]. Suresh et al. Memory augmented recurrent neural networks for de novo drug design. PLoS One. 2022. https://doi.org/10.1371/journal.pone.0269461
[65]. Mao et al. Transformer-based molecular generative model for antiviral drug design. J Chem Inf Model. 2023. https://doi.org/10.1021/acs.jcim.3c00536
[66]. Bess et al. Artificial intelligence to discover antimicrobial agents for infectious diseases. Drug Discov Today. 2022. https://doi.org/10.1016/j.drudis.2021.10.022
[67]. Tripathi et al. Generative adversarial networks in drug discovery and targeting. Artif Intell Life Sci. 2022. https://doi.org/10.1016/j.ailsci.2022.100045
[68]. Rafiei et al. DeepTraSynergy: drug combinations using multimodal transformer learning. Bioinformatics. 2023. https://doi.org/10.1093/bioinformatics/btad438
[69]. Shimizu et al. LIGHTHOUSE illuminates therapeutics for diverse diseases. iScience. 2022. https://doi.org/10.1016/j.isci.2022.105314
[70]. Tang et al. AI-aided design of targeted covalent inhibitors against SARS-CoV-2. Biomolecules. 2022. https://doi.org/10.3390/biom12060746
[71]. Arora et al. Artificial intelligence: a virtual chemist for natural product drug discovery. J Biomol Struct Dyn. 2024. https://doi.org/10.1080/07391102.2023.2216295
[72]. Othman et al. Artificial intelligence for natural product drug discovery and development: current landscape, applications, and future directions. Intell Based Med. 2025. https://doi.org/10.1016/j.ibmed.2025.100316
[73]. Rutz et al. The LOTUS initiative for open knowledge management in natural products research. eLife. 2022. https://doi.org/10.7554/eLife.70780
[74]. Liao et al. Molecular insights fast-tracked: AI in biosynthetic pathway research. Nat Prod Rep. 2025. https://doi.org/10.1039/d4np00003j
[75]. Zhang et al. Machine learning approaches for elucidating the biological effects of natural products. Nat Prod Rep. 2021. https://doi.org/10.1039/d0np00043d
[76]. Brittin et al. Machine learning-based bioactivity classification of natural products using LC-MS/MS metabolomics. J Nat Prod. 2025. https://doi.org/10.1021/acs.jnatprod.4c01123
[77]. Xiong et al. ADMETlab 2.0: an integrated online platform for ADMET prediction. Nucleic Acids Res. 2021. https://doi.org/10.1093/nar/gkab255
[78]. Swanson et al. ADMET-AI: a machine learning ADMET platform. Bioinformatics. 2024. https://doi.org/10.1093/bioinformatics/btae416
[79]. Qadri et al. Explainable artificial intelligence in drug discovery and development. Pharmaceutics. 2025. https://doi.org/10.3390/pharmaceutics17091119
[80]. Lavecchia et al. Explainable artificial intelligence approaches in drug discovery. WIREs Comput Mol Sci. 2025. https://doi.org/10.1002/wcms.70049
[81]. Chihomvu et al. AI in phytochemical research and medicinal plant drug discovery. Int J Mol Sci. 2024. https://doi.org/10.3390/ijms25168792
[82]. Nothias-Esposito et al. Molecular networking in Euphorbia natural product chemical analysis. J Nat Prod. 2024. https://doi.org/10.1021/acs.jnatprod.4c00190
[83]. Berger et al. MassQL and molecular networking in natural products chemistry. J Nat Prod. 2024. https://doi.org/10.1021/acs.jnatprod.3c00750
[84]. Sheng et al. Integrated molecular networking for natural product discovery. Anal Chem. 2024. https://doi.org/10.1021/acs.analchem.3c04746
[85]. Rios-Martinez et al. Self-supervised learning for biosynthetic gene cluster analysis. PLoS Comput Biol. 2023. https://doi.org/10.1371/journal.pcbi.1011162
[86]. Yuan et al. Machine learning-enabled genome mining and bioactivity prediction of natural products. ACS Synth Biol. 2023. https://doi.org/10.1021/acssynbio.3c00234

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

DOI: 10.17148/IARJSET.2026.131122

[87]. Wang et al. Molecular networking as an efficient tool for natural product research. J Pharm Biomed Anal. 2025. https://doi.org/10.1016/j.jpba.2025.116741

[88]. Li et al. Database-assisted molecular networking in natural products. J Chromatogr A. 2023. https://doi.org/10.1016/j.chroma.2023.464417

[89]. Seshadri et al. Synthetic biology in natural product biosynthesis. Chem Rev. 2025. https://doi.org/10.1021/acs.chemrev.4c00567

[90]. Basnet et al. Advances in AI-based strategies and tools to facilitate natural product and drug development. Crit Rev Biotechnol. 2025. https://doi.org/10.1080/07388551.2025.2478094

[91]. Catacutan et al. Machine learning in preclinical drug discovery. Nat Chem Biol. 2024. https://doi.org/10.1038/s41589-024-01679-1

[92]. Singh et al. Natural products as drug leads: exploring their potential in drug discovery and development. Naunyn Schmiedebergs Arch Pharmacol. 2025. https://doi.org/10.1007/s00210-024-03622-6

[93]. Li et al. Automatic MS/MS data mining strategy for discovering target natural products. Anal Chem. 2022. https://doi.org/10.1021/acs.analchem.2c01559

[94]. Jeon et al. Predicting biochemical and physiological effects of natural products from molecular structures using machine learning. Nat Prod Rep. 2021. https://doi.org/10.1039/d1np00016k

[95]. Gangwal et al. Current strategies to address data scarcity in artificial intelligence-based drug discovery. Comput Biol Med. 2024. https://doi.org/10.1016/j.compbiomed.2024.108734

[96]. Dührkop et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. Nat Biotechnol. 2021. https://doi.org/10.1038/s41587-020-0740-8

[97]. Huber et al. MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra. J Cheminform. 2021. https://doi.org/10.1186/s13321-021-00558-4

[98]. Huber et al. Spec2Vec: improved mass spectral similarity scoring through learning of structural relationships. PLoS Comput Biol. 2021. https://doi.org/10.1371/journal.pcbi.1008724

[99]. Öztürk et al. Exploring chemical space using natural language processing methodologies for drug discovery. Drug Discov Today. 2020. https://doi.org/10.1016/j.drudis.2020.01.020

[100]. Wu et al. Transfer learning with molecular graph convolutional networks for GPCR ligand bioactivity prediction. Comput Biol Chem. 2022. https://doi.org/10.1016/j.compbiolchem.2022.107664

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License