IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

Artificial Intelligence in Natural Products Discovery

Gaurav S. Mitkari[1], Trupti Y. Biradar[2], Aditya K. Tanpure[3], Mr. Vikas Mugle[4], Dr. V. M. Dharashive[5]
[1],[2],[3] Students, [4] Guide, [5] Principal
Shivlingshwar College of Pharmacy, Almala
[1] mitkarigaurav64@gmail.com

Abstract: *Natural products (NPs) remain a valuable source of chemical diversity for pharmacological research and drug development. Their complex structures and biological activities have contributed to the discovery of medicines and inspired new therapeutic strategies. Nevertheless, conventional natural-product discovery can be slow and resource-intensive because researchers must address source selection, extraction, isolation, dereplication, structural elucidation, biological testing and subsequent optimization. Artificial intelligence (AI), including machine learning (ML), deep learning (DL), natural-language processing, computer vision and graph-based methods, offers computational approaches to help manage these tasks. This review discusses the role of AI across the natural-product discovery pipeline, with emphasis on data integration, plant and microbial source exploration, biosynthetic gene-cluster analysis, metabolomics and spectral interpretation, activity and target prediction, virtual screening, network pharmacology, toxicity and ADMET estimation, and molecular design. It also describes commonly used databases and computational resources, outlines a practical AI-supported workflow, and considers opportunities for integrating experimental evidence with computational predictions. The benefits of AI include prioritization of candidate compounds, improved analysis of large and heterogeneous datasets, hypothesis generation and more systematic exploration of chemical and biological relationships. Important limitations remain, including dataset bias, inconsistent annotation, small sample sizes, overfitting, poor external validation, limited interpretability, reproducibility concerns and the need for experimental confirmation. Ethical, regulatory and biodiversity considerations are also discussed. AI is best regarded as a decision-support layer that complements not replaces phytochemistry, pharmacology, analytical chemistry and biological experimentation. A responsible integration of AI with laboratory validation and transparent reporting may strengthen the discovery and development of natural-product-inspired therapeutic candidates..*

Keywords: Artificial intelligence; natural products; machine learning; drug discovery; metabolomics; genome mining; phytochemicals; virtual screening; bioinformatics.

I. INTRODUCTION

Natural products are molecules produced by living organisms, including plants, fungi, bacteria, marine organisms and animals. They encompass a broad range of chemical classes such as alkaloids, terpenoids, flavonoids, polyketides, peptides, phenolics and glycosides. Natural products and their derivatives have historically contributed to pharmacotherapy and continue to provide chemical scaffolds for research. Their structures often contain stereochemistry, functional groups and ring systems that differ from many conventional screening-library molecules, making them valuable sources of chemical inspiration [1 3].

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-41057

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, November 2026 Impact Factor: 8.2

Despite their promise, discovery from natural sources is frequently complicated by sample complexity, low abundance, seasonal and geographical variation, uncertainty in identification, repeated isolation of known compounds, and difficulties connecting a chemical feature to a biological effect. Traditional approaches often combine ethnobotanical information, extraction, fractionation, bioassays, purification and structure determination. These steps remain essential, but they may consume considerable time and resources, particularly when the initial candidate space is large [2,4].

AI methods are increasingly applied to biomedical and chemical datasets. Algorithms can learn relationships among chemical structures, biological targets, activity measurements, spectral data, sequences and textual records. In natural-product research, these capabilities may assist prioritization and pattern recognition in datasets that are too complex for manual interpretation alone. The most useful implementation is typically an integrated workflow in which predictions guide experimental choices and new results are used to refine subsequent analyses [5,6].

This review provides an overview of AI applications in natural-product discovery and development, summarizes major approaches and resources, and critically discusses limitations and future directions. The review is intended as a broad, educational synthesis rather than a systematic review or meta-analysis; the selection and organization of sources should be updated and checked before submission to a journal.

II. REVIEW APPROACH AND SCOPE

This narrative review organizes published research and review literature around the stages of natural-product discovery where AI may provide assistance. The themes include data resources, source and biosynthetic exploration, metabolite annotation, molecular representation, bioactivity and target prediction, screening, network pharmacology, ADMET and toxicity estimation, generative design and validation. The discussion draws on relevant literature identified through scholarly search and reference-chain exploration. It does not claim exhaustive coverage, formal risk-of-bias assessment or systematic database searching. For a journal submission, authors should document the databases, search dates, search strings, inclusion criteria and screening process, particularly if the journal expects a systematic or scoping review.

III. NATURAL PRODUCTS IN DRUG DISCOVERY

Natural products have been investigated across antibacterial, antifungal, anticancer, anti-inflammatory, cardiovascular, neurological and other therapeutic areas. Their biological effects may arise from interactions with a single molecular target, multiple targets, pathways or cellular processes. Some natural-product-derived medicines are chemically modified or semi-synthetic, reflecting the use of natural scaffolds as starting points rather than necessarily as final drug substances [1,3].

Natural-product research includes several source categories. Plant metabolites are commonly explored through pharmacognosy and phytochemistry. Microorganisms are studied through fermentation, genomic analysis and biosynthetic pathways. Marine organisms and other ecological sources may yield unusual structures, while synthetic biology can enable production or modification of compounds that are difficult to obtain from their original sources. AI can potentially connect these different source types to chemical, biological and environmental datasets.

3.1 Traditional discovery workflow

A typical discovery workflow begins with source selection and collection, followed by authentication or characterization of the source. Researchers then prepare extracts, conduct preliminary chemical or biological screening, fractionate samples, isolate compounds, identify structures and test biological activity. Repeated compounds may be recognized through dereplication. Candidate compounds may undergo further characterization, selectivity testing, mechanism-of-action studies and pharmacokinetic or safety evaluation. Each stage can generate data that may inform the next stage.

3.2 Bottlenecks and opportunities

- Complex mixtures and difficult separation of individual constituents.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

- Rediscovery of known compounds when dereplication is incomplete.
- Limited availability of authenticated reference materials and high-quality datasets.
- Uncertain target identity and mechanism for compounds with complex activity profiles.
- Variation in sample preparation, source provenance, assay conditions and reporting.
- Challenges in translating in-vitro activity into exposure, safety and efficacy in living systems.

AI may help address some bottlenecks by integrating datasets and prioritizing candidates, but the quality of the underlying data and experimental design remains a major determinant of useful output [5,7].

IV. ARTIFICIAL INTELLIGENCE: CONCEPTS AND METHODS

AI is a broad term for computational approaches designed to perform tasks associated with reasoning, pattern recognition, prediction or decision support. ML is a subset of AI in which algorithms learn statistical relationships from data. DL is a subset of ML that uses multi-layer neural networks to represent complex patterns. Natural-product applications may also use NLP for literature extraction, computer vision for images and graph algorithms for chemical or biological networks [5,8].

4.1 Machine learning

Supervised learning uses labeled examples to learn relationships between input features and outcomes, such as chemical descriptors and measured activity. Classification models predict categories, while regression models estimate continuous values. Unsupervised methods identify patterns without predefined labels, including clustering, dimensionality reduction and anomaly detection. Semi-supervised and self-supervised approaches can use unlabeled data alongside limited labeled examples, although their usefulness depends on data quality and the relevance of learned representations.

4.2 Deep learning

Deep-learning architectures include convolutional neural networks, recurrent networks, transformers and graph neural networks. Neural models may be applied to sequences, images, spectra, text or molecular graphs. Their capacity to model complex relationships is useful in some contexts, but can also increase the risk of overfitting when datasets are small. Appropriate validation, regularization and reporting are therefore important [5,9].

4.3 Natural-language processing and large language models

NLP methods can extract entities and relationships from scientific literature, including species names, compounds, assay descriptions, target names and experimental outcomes. Large language models may assist literature organization, summarization, data-entry support and code drafting. However, generated outputs may contain unsupported claims or fabricated references. Human review, source verification, privacy awareness and careful handling of copyrighted material are necessary. NLP extraction should preserve the original context, units, assay details and uncertainty rather than treating every extracted statement as an equally reliable fact.

4.4 Molecular representation

Molecules may be represented by molecular descriptors, fingerprints, SMILES strings, three-dimensional coordinates or graphs. Descriptors summarize properties such as size, polarity and connectivity. Fingerprints encode structural fragments or patterns. Graph representations describe atoms as nodes and bonds as edges, and graph neural networks can learn representations from these structures. The choice of representation should match the question, dataset size and validation strategy [5].

V. AI ACROSS THE NATURAL-PRODUCT DISCOVERY PIPELINE

5.1 Source selection and biodiversity-informed discovery

AI can integrate information about organisms, habitats, taxonomic relationships, chemical reports, ecological interactions and prior screening outcomes to help prioritize research questions. Text-mining approaches may organize ethnobotanical or historical information, while data-driven approaches may identify associations between source taxa and reported chemical classes. Such associations should not be interpreted as proof of efficacy or safety. Source

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, November 2026 Impact Factor: 8.2

selection should follow ethical collection, conservation, access-and-benefit-sharing requirements and appropriate permissions.

5.2 Biosynthetic gene-cluster mining

Microbial genomes encode biosynthetic gene clusters (BGCs) that may be responsible for the production of specialized metabolites. Genome-mining software can identify and annotate candidate clusters, compare cluster organization and estimate the likely product class. ML and DL approaches may improve prioritization, especially when combined with sequence, domain, phylogenetic and metabolomic information. Tools such as antiSMASH and curated resources such as MIBiG have contributed to computational biosynthetic research [10 12]. Predictions are hypotheses and require appropriate biological and chemical confirmation.

5.3 Metabolomics and mass-spectrometry analysis

Metabolomics can generate large datasets from chromatography, mass spectrometry and tandem-MS experiments. AI methods may support peak detection, feature grouping, spectral similarity, molecular formula or candidate annotation, and molecular networking. Spectral-learning approaches can identify patterns in fragmentation data and help organize related compounds. Annotation confidence should be reported using clear evidence levels, and tentative identifications should not be presented as confirmed structures [13 15].

5.4 Structure elucidation and dereplication

Dereplication seeks to recognize previously reported compounds early in the workflow, reducing repeated isolation and characterization. Computational tools can compare mass spectra, molecular formulae, predicted fragments, spectral libraries and structural databases. AI-supported spectral interpretation may suggest candidate structures or related molecular families, while NMR data and complementary evidence can assist structure elucidation. A prediction alone is not a definitive structural assignment; authentic standards, orthogonal analytical evidence and expert interpretation may be required [13 15].

5.5 Biological activity prediction

Activity prediction models may estimate whether a molecule is likely to exhibit a particular biological effect or fall within a defined assay range. Training data may include curated activity measurements, molecular structures, assay metadata and target information. Such models can prioritize candidates for follow-up, but reported activity values are affected by assay conditions, measurement quality, compound purity and dataset selection. Models should be evaluated using data partitions that reduce leakage and reflect the intended use.

5.6 Target identification and mechanism-of-action hypotheses

AI-supported target prediction can combine chemical similarity, protein sequences, target structures, interaction data, gene-expression patterns and pathway information. Network methods may connect compounds to proteins, pathways or phenotypes, generating hypotheses about mechanisms. The resulting associations may be useful for designing follow-up experiments. Computational target predictions should be distinguished from direct target-engagement evidence and mechanistic confirmation.

5.7 Virtual screening and molecular docking

Virtual screening can prioritize compounds from collections using structure-based or ligand-based approaches. Docking estimates possible binding poses and scores interactions in a target structure, while ML models may rerank candidates or estimate binding-related properties. Natural products can pose additional challenges because of stereochemical complexity, flexibility, unusual functional groups and incomplete structural data. Docking scores are not experimental binding measurements, and reliable screening requires appropriate controls, uncertainty analysis and biological validation [5,16].

5.8 Network pharmacology

Network pharmacology models relationships among compounds, targets, pathways, disease-associated genes and phenotypic observations. Natural-product mixtures and multi-component preparations are sometimes investigated using network approaches because their effects may involve multiple components and biological systems. Network predictions can help formulate hypotheses, but database-derived links may be incomplete, biased or indirect.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, November 2026 Impact Factor: 8.2

Experimental tests are needed to distinguish causal effects from correlations and to evaluate the contribution of individual components.

5.9 ADMET and safety-related prediction

Computational models may estimate properties related to absorption, distribution, metabolism, excretion and toxicity. Examples include physicochemical descriptors, permeability-related estimates, metabolic stability, potential interaction alerts and toxicity endpoints. Prediction is useful for prioritization and early risk awareness, not as a substitute for toxicology, pharmacokinetic studies or regulatory evaluation. Models trained on synthetic compounds may perform differently on natural products, and domain applicability should be assessed carefully.

5.10 Generative AI and natural-product-inspired design

Generative models can propose molecular structures or modify existing representations according to learned patterns and design constraints. Natural-product-inspired design may explore scaffold variations, molecular properties and structural novelty. Generative output is not automatically synthesizable, safe, active or environmentally responsible. Candidate generation should be constrained by chemical validity, feasibility, evidence, novelty, intellectual-property considerations and the need for appropriate laboratory evaluation. The role of generative AI is to propose hypotheses not to certify candidate molecules [5,17].

VI. DATA RESOURCES AND COMPUTATIONAL ECOSYSTEM

Natural-product discovery relies on several complementary resource types: chemical structure databases, spectral libraries, biosynthetic repositories, genomic databases, protein and pathway resources, bioactivity databases and scientific literature. Examples include COCONUT for natural-product structures, GNPS for mass-spectrometry-based molecular networking and spectral data, MIBiG for experimentally characterized biosynthetic gene clusters, and antiSMASH for genome-mining analysis [10 12,14]. These resources have different coverage and update cycles, so investigators should record version, date of access and data provenance.

6.1 Data quality and harmonization

AI-ready datasets require standardized identifiers, consistent units, assay descriptions, provenance, source taxonomy, extraction and sample details, structure representations and confidence labels. Duplicate records should be identified; conflicts should be preserved or resolved transparently. Negative results and uncertainty are valuable when collected and reported appropriately. Dataset harmonization should not erase important differences between experimental methods or biological contexts.

6.2 Data bias and coverage

Data availability is not uniform across organisms, regions, compound classes, therapeutic areas or assays. Highly studied taxa and molecules may dominate datasets, while poorly characterized biodiversity and unpublished findings remain underrepresented. This can affect model performance and generalizability. Researchers should examine the composition of training and test datasets and describe limitations in coverage.

VII. PRACTICAL AI-SUPPORTED WORKFLOW

A responsible computational workflow can be organized into iterative steps. First, define a research question and intended prediction. Second, assemble data with clear provenance and quality criteria. Third, clean and harmonize records while retaining relevant metadata. Fourth, select a representation and baseline model. Fifth, partition data using a strategy suited to the problem, including scaffold, temporal or source-aware splits where appropriate. Sixth, train and compare models. Seventh, evaluate using suitable metrics and external data when possible. Eighth, estimate uncertainty and applicability. Ninth, prioritize candidates for independent experimental assessment. Finally, incorporate experimental results into a revised dataset and workflow.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

7.1 Workflow diagram (text format)

Research question → Data acquisition → Data curation and harmonization → Molecular / genomic / spectral representation → Model selection and training → Internal and external validation → Uncertainty and applicability assessment → Candidate prioritization → Experimental confirmation → Data feedback and model refinement

7.2 Model evaluation

Evaluation metrics should match the task. Classification may use precision, recall, specificity, F1-score, area under the ROC curve and calibration. Regression may use mean absolute error, root mean squared error and calibration analysis. For imbalanced datasets, precision recall analysis may be informative. Model evaluation should include confidence intervals when feasible, suitable baselines, external validation and a clear description of the data partition. Random splits alone may produce overly optimistic results if structurally or experimentally related records occur in both training and test sets.

7.3 Reproducibility

Reproducible reporting should include dataset sources and dates, inclusion criteria, preprocessing steps, model architecture or algorithm, hyperparameters, software versions, evaluation protocol, random seeds where relevant, metrics, limitations and code or data availability statements. Computational notebooks should be reviewed for correctness and should not expose confidential or sensitive data.

VIII. APPLICATIONS AND REPRESENTATIVE RESEARCH THEMES

8.1 Antimicrobial discovery

Natural-product research has historically contributed to antimicrobial discovery. AI may assist in organizing microbial genomic information, identifying candidate biosynthetic clusters, predicting molecular properties or prioritizing compound collections for screening. Research reports should distinguish computational predictions from experimentally confirmed antimicrobial activity and should document the biological assay, organism, controls and reproducibility.

8.2 Anticancer research

AI methods may integrate chemical, molecular, cellular and omics data to prioritize natural-product candidates for oncology research. Phenotypic screening, target prediction and pathway analysis can generate hypotheses about activity. However, cell-based effects may reflect general cytotoxicity, assay interference or non-specific mechanisms. Selectivity, mechanistic evidence and appropriately designed validation are necessary before translational claims are made.

8.3 Anti-inflammatory and immunomodulatory research

Natural products are frequently investigated for effects on inflammatory signaling and immune-related processes. AI can analyze pathway, gene-expression and activity data to identify patterns and propose candidate mechanisms. The interpretation of such results requires careful consideration of biological context, experimental model limitations and the difference between pathway association and demonstrated causality.

8.4 Metabolic and neurological research

In metabolic and neurological research, AI can support the integration of chemical, phenotypic and omics datasets. Candidate prioritization may involve target and pathway hypotheses, property prediction and literature organization. Biological complexity, model-system differences and the challenges of exposure at relevant tissues remain important considerations.

8.5 Natural-product mixtures and herbal preparations

Multi-component preparations introduce challenges in composition, interactions, batch consistency and attribution of effects. Network pharmacology and data-integration approaches can help organize hypotheses about components and pathways. Nevertheless, mixture-level predictions are particularly sensitive to incomplete composition data, uncertain exposure and database assumptions. Analytical characterization and controlled experimental evaluation are essential.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

IX. BENEFITS AND OPPORTUNITIES

- **Prioritization:** helps reduce the number of candidates requiring intensive experimental investigation.
- **Data integration:** connects chemical, genomic, spectral, biological and literature data.
- **Pattern discovery:** identifies relationships that may not be apparent through manual analysis.
- **Hypothesis generation:** supports candidate targets, pathways and mechanisms for further study.
- **Efficiency:** may improve planning and allocation of screening or analytical resources.
- **Iterative learning:** experimental data can refine computational models and future prioritization.
- **Collaboration:** encourages integration among pharmacognosy, chemistry, bioinformatics, pharmacology and data science.

These potential benefits are context-dependent. A model that improves ranking in one dataset may not improve the success rate, cost or time of a real discovery program. Prospective studies and transparent comparisons with traditional approaches are needed to establish practical impact [5,6].

X. LIMITATIONS AND CHALLENGES

10.1 Small and noisy datasets

Natural-product datasets may contain few confirmed examples, heterogeneous assays and incomplete records. Small datasets can lead to unstable models and overfitting. Noise may arise from differences in extraction, compound purity, assay protocols, measurement units and reporting. Carefully curated data and simple baseline methods may sometimes outperform complex architectures.

10.2 Validation and data leakage

Data leakage occurs when information related to the test set enters model development, producing inflated performance estimates. Similar compounds, replicate measurements, related assays or records from the same study may appear across partitions. Validation design should reflect the intended application and, where possible, use scaffold, time, study or source-based separation. External validation is particularly important.

10.3 Interpretability and uncertainty

Complex models may produce predictions that are difficult to explain. Explainability techniques can suggest which features contribute to a prediction, but they do not necessarily establish causal mechanisms. Uncertainty estimates and applicability-domain checks can help identify predictions that should not be trusted. Confidence should be reported rather than implied.

10.4 Chemical and biological complexity

Natural products may have unusual scaffolds, stereochemistry, flexibility or chemical instability. Biological activity may depend on metabolites, formulation, context or interactions with multiple pathways. A computational representation may omit relevant features, and model performance may not transfer to chemically unusual or poorly represented compounds.

10.5 Experimental confirmation

AI output requires validation using appropriate analytical and biological methods. Confirmatory work may include structure verification, purity assessment, reproducible assays, selectivity evaluation, target engagement, mechanism studies and pharmacokinetic or safety investigations as appropriate to the research stage. The required evidence depends on the claim being made.

10.6 Cost, skills and infrastructure

Data infrastructure, computational resources and specialist expertise can be barriers. Researchers need training in dataset design, statistical analysis, programming, scientific interpretation and responsible reporting. Interdisciplinary collaboration and open resources may help, but tool availability alone does not guarantee reliable results.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

XI. ETHICAL, LEGAL AND SUSTAINABILITY CONSIDERATIONS

Natural-product research may involve biodiversity, traditional knowledge, community rights and access-and-benefit-sharing obligations. Collection and use of biological materials should comply with applicable laws, institutional policies, permits and ethical requirements. AI systems may reproduce biases from historical research patterns, including unequal representation of species, regions or communities. Data provenance and appropriate recognition of contributors are important.

Researchers should also consider intellectual-property rights, software and database licences, privacy where human-derived data are involved, and transparency in AI-assisted writing or analysis. Environmental considerations include sustainable sourcing, conservation, waste management and the possibility of biosynthetic or other approaches that reduce pressure on natural populations. Claims about sustainability should be supported rather than assumed.

XII. FUTURE DIRECTIONS

Future work may benefit from multimodal models that combine molecular structures, spectra, sequence data, biological observations and literature. Improved annotation standards and curated datasets could strengthen model training and evaluation. Foundation models may provide reusable representations, but their performance and reliability in natural-product domains require independent assessment. Experimental feedback loops and automation may improve throughput in specific settings, provided data quality and reproducibility remain central.

The field may also move toward more prospective evaluations, where candidate predictions are registered or documented before experimental testing. Such designs can help measure whether AI improves discovery outcomes rather than simply reproducing known associations. Better reporting of negative findings, uncertainty and model failures would support more realistic assessment. Translation will require coordinated efforts across chemistry, biology, data science, pharmacology, manufacturing, regulation and conservation.

XIII. CONCLUSION

AI offers a collection of computational methods that may support natural-product discovery at multiple stages, from source and biosynthetic exploration to spectral annotation, activity prediction, target hypotheses, virtual screening and property estimation. The most promising use is not a fully automated replacement for laboratory research but an integrated and iterative approach that helps researchers organize evidence and prioritize questions. Persistent challenges include data quality, bias, validation, interpretability, chemical and biological complexity, reproducibility and experimental confirmation. Transparent workflows, well-curated datasets, responsible biodiversity practices and prospective validation will be essential for translating computational predictions into meaningful scientific advances.

ACKNOWLEDGEMENTS

The authors may add acknowledgements for institutional support, funding, mentorship or contributions, as appropriate.

Funding

No funding statement should be included unless accurate. Insert the actual funding information or state 'No external funding was received' only after confirming this statement.

CONFLICT OF INTEREST

The authors should disclose any actual competing interests. If none exist, use the journal's approved wording after confirming.

AUTHOR CONTRIBUTIONS

Insert a contribution statement using the journal's preferred taxonomy and only after confirming each author's contribution.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

DATA AVAILABILITY

No new datasets were generated for this narrative review. This statement should be revised if the authors generate or use additional datasets.

REFERENCES

1. Mullowney MW, Duncan KR, Elsayed SS, et al. Artificial intelligence for natural product drug discovery. Nat Rev Drug Discov. 2023;22:895 916. doi:10.1038/s41573-023-00774-7.
2. Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod. 2020;83:770 803.
3. Atanasov AG, Zotchev SB, Dirsch VM, et al. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov. 2021;20:200 216.
4. Harvey AL, Edrada-Ebel R, Quinn RJ. The re-emergence of natural products for drug discovery in the 21st century. Nat Rev Drug Discov. 2015;14:111 129.
5. Skinnider MA, et al. Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. Nat Commun. 2020;11:6058.
6. Medema MH, Fischbach MA. Computational approaches to natural product discovery. Nat Chem Biol. 2015;11:639 648.
7. Sahayasheela VJ, Lankadasari MB, Dan VM, et al. Artificial intelligence in microbial natural product drug discovery: current and emerging role. Nat Prod Rep. 2022;39:2215 2230. doi:10.1039/D2NP00035K.
8. Pant A, Jena GK, Ezzat SM, et al. Roles of artificial intelligence and machine learning approach in natural products-based drug discovery. In: Phytochemistry, Computational Tools and Databases in Drug Discovery. Elsevier; 2023.
9. Huber F, et al. Spec2Vec: improved mass spectral similarity scoring through learning of structural relationships. PLoS Comput Biol. 2021;17:e1008724.
10. Blin K, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021;49:W29 W35.
11. Medema MH, de Rond T, Moore BS. Mining genomes to illuminate the specialized chemistry of life. Nat Rev Genet. 2021;22:553 571.
12. Terlouw BR, et al. MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. Nucleic Acids Res. 2023;51:D603 D610.
13. Huber F, van der Burg S, van der Hooft JJJ, Ridder L. MS2DeepScore: a novel deep learning similarity measure to compare tandem mass spectra. J Cheminform. 2021;13:84.
14. Ludwig M, et al. Database-independent molecular formula annotation using Gibbs sampling through ZODIAC. Nat Mach Intell. 2020;2:629 641.
15. Hoffmann MA, et al. High-confidence structural annotation of metabolites absent from spectral libraries. Nat Biotechnol. 2022;40:411 422.
16. Choudhury C, Arul Murugan N, Priyakumar UD. Structure-based drug repurposing: traditional and advanced AI/ML-aided methods. Drug Discov Today. 2022;27:1847 1861.
17. Liu X, IJzerman AP, van Westen GJP. Computational approaches for de novo drug design: past, present, and future. Methods Mol Biol. 2021;2190:139 165.
18. van der Hooft JJJ, et al. Linking genomics and metabolomics to chart specialized metabolic diversity. Chem Soc Rev. 2020;49:3297 3314.
19. Newman DJ, Cragg GM. Natural products as sources of new drugs: a continuing story. (Authors should verify edition/year and exact citation details before submission; this placeholder should be replaced by the appropriate source if used.)
20. AI-driven drug discovery from natural products. Advanced Agrochem. 2024;3(3):185 187.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, November 2026
Impact Factor: 8.2

doi:10.1016/j.aac.2024.06.003.

Suggested graphical abstract and figures (to prepare separately)

Figure 1. AI-integrated natural-product discovery pipeline: source selection, genomic and metabolomic data, AI analysis, candidate prioritization, laboratory validation and iterative feedback.

Figure 2. AI methods mapped to natural-product data types: sequence models for biosynthetic data; spectral models for MS/MS; molecular graphs for chemical structure; NLP for literature; network methods for compound target pathway relationships.

Figure 3. Challenges and enabling factors: data quality, annotation, validation, interpretability, biodiversity stewardship, experimental confirmation and reproducibility.

Before journal submission: author checklist

- Verify every reference against the original article, DOI and journal format; correct any citation errors or placeholders.
- Update the literature search and include recent relevant papers without duplicating existing reviews.
- Select a target journal and adapt the scope, length, reference style, figure requirements and declarations.
- Add original, properly licensed figures or redraw figures with captions and source attribution; do not copy publisher figures without permission.
- Have all authors critically review the manuscript, confirm authorship and disclosures, and check factual accuracy.
- Run similarity checking using an institutional or publisher-approved tool; revise overlapping phrasing and cite sources appropriately.
- Do not claim that AI-generated text is guaranteed to be plagiarism-free or undetectable. Follow the journal's AI-use disclosure policy.