Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.53555/zhnsna60

ARTIFICIAL INTELLIGENCE IN PHYTOCHEMICAL RESEARCH: MAPPING THE PLANT KINGDOM FOR DRUG DISCOVER

Harikrishna Ramaprasad Saripalli[1*], Rajasekhar Dega[2], Uma Devarapalli[3], K. Aruna Kumari[3], Nitchal Kiran Jaladi[4], P.V. Hemalatha[5], Jala Aaron Hemanth Reuveng[6]

[1*]Technical Director, King Honey Limited, Owned and Operated by Me|Today Limited, 1/6 Mahoe Street, Tauhara, RD2, Taupo 3378, New Zealand.
[2]Lecturer in Botany, K.R.K Government Degree College, Addanki, Andhra Pradesh, India.
[3]Lecturer in Botany, Government College for Women (Autonomous), Guntur, Andhra Pradesh 522002, India.
[4]Professor, Department of Physics, Vignan's Foundation for Science, Technology and Research, Deemed to be University, Guntur, Andhra Pradesh, 522213, India.
[5]Senior Lecturer in Chemistry, K.R.K Government Degree College, Addanki, Andhra Pradesh, India.
[6]B.Tech. Student, Department of Biotechnology, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Andhra Pradesh, Guntur, 522502, India.

***Corresponding Author:** Harikrishna Ramaprasad Saripalli
*Email: haris@kinghoney.co.nz

ABSTRACT

The vast diversity of the plant kingdom represents both an incredible reservoir of potential therapeutic agents and a formidable challenge for drug discovery. Traditional approaches to identifying medicinally valuable plants are labor-intensive, time-consuming, and often limited by narrow ethnobotanical scope. In recent years, the integration of artificial intelligence (AI) and machine learning (ML) has revolutionized this process, offering new methods to analyze and prioritize botanical species with pharmacological potential more efficiently and systematically.

This review explores the transformative role of AI and ML in modern phytochemical research and natural product discovery. It highlights how these technologies are enabling more precise, data-driven selection of plant species by leveraging diverse datasets—including ethnobotanical records, phytochemical compositions, and bioactivity profiles. AI-powered predictive models can identify patterns and correlations across complex datasets that may not be apparent through conventional statistical methods. This leads to the early detection of bioactive compounds and novel therapeutic leads.

Advancements in deep learning and computer vision have also enhanced image-based plant identification and disease diagnostics, improving taxonomic accuracy and aiding in species classification, especially in under-documented regions. Moreover, the integration of omics data—genomics, transcriptomics, proteomics, and metabolomics—with AI models is facilitating a systems-level understanding of plant metabolic pathways and their therapeutic outputs.

Despite its promise, AI-driven botanical research faces challenges, including data quality, model interpretability, and the need for interdisciplinary collaboration. Nonetheless, the potential for accelerating natural product discovery through AI is profound. By automating and optimizing the

exploration of plant biodiversity, AI stands to unlock new frontiers in drug discovery, conservation biology, and personalized medicine.
This review presents a comprehensive overview of the current landscape, identifies key breakthroughs and limitations, and outlines future directions for AI-assisted exploration of the botanical seascape.

Keywords: Artificial intelligence, machine learning, botanical therapeutics, phytochemical research, natural product discovery, deep learning, computer vision, omics integration, plant species identification, bioactive compounds, ethnobotany, drug discovery, metabolomics, proteomics, genomics, AI interpretability, interdisciplinary collaboration

INTRODUCTION

The plant kingdom comprises an immense diversity of species—estimated in the hundreds of thousands—many of which possess untapped therapeutic potential. Historically, the identification of medicinal plants has relied heavily on ethnobotanical knowledge and bioassay-guided fractionation, a process that, while valuable, is often slow, labor-intensive, and limited in scale. This traditional approach struggles to keep pace with the urgent demand for novel bioactive compounds, especially in the face of emerging health challenges and drug-resistant pathogens.
Artificial intelligence (AI) and machine learning (ML) are emerging as transformative tools in this domain. These technologies have the capacity to integrate and analyze complex, multidimensional data from a variety of sources, including ethnobotanical records, phytochemical and genomic databases, and experimental bioactivity assays. By learning from patterns within this data, AI/ML models can rapidly triage and prioritize plant species with high therapeutic potential, significantly streamlining the discovery pipeline. This integration not only improves efficiency but also enhances the likelihood of identifying novel compounds with clinically relevant bioactivity.

Ethnobotany Meets Machine Learning

Recent research has demonstrated that artificial intelligence (AI) models can significantly outperform traditional ethnobotanical heuristics in identifying plants with medicinal potential. For example, in a large-scale study involving approximately 21,100 species across the plant families Apocynaceae, Loganiaceae, and Rubiaceae, machine learning (ML) classifiers—including support vector machines (SVM), logistic regression, gradient boosting, and Bayesian neural networks—achieved a predictive precision of approximately 67% for identifying species with antiplasmodial activity. In contrast, species selected using conventional ethnobotanical knowledge alone yielded a lower precision of about 46% (Richard-Bollans et al., 2023). These findings highlight the ability of ML algorithms to detect subtle, multidimensional patterns in large datasets, uncovering therapeutic potential that might otherwise be overlooked.
However, while AI offers superior scalability and analytical power, indigenous and traditional knowledge remains a critical component of the discovery process. Ethnobotanical insights provide contextual understanding of plant use, ecological interactions, and cultural significance offering valuable hypotheses that can guide and inform data-driven exploration. The integration of AI with human-informed expertise thus represents a synergistic model for accelerating drug discovery from plants.

AI in Phytochemical & Omics Profiling

Artificial intelligence (AI) is increasingly being leveraged to analyse large-scale omics datasets—including genomics, transcriptomics, proteomics, and metabolomics—to identify and characterize bioactive compounds within plant systems. These high-dimensional datasets offer detailed molecular insights, but their complexity often makes them difficult to interpret using traditional bioinformatics tools. AI and machine learning (ML) algorithms, particularly deep learning models,

have proven highly effective in extracting meaningful patterns and correlations from such data, enabling the discovery of novel metabolites with therapeutic relevance (Varghese et al., 2025).
Applications of AI in this context include metabolite identification, structural elucidation, and the mapping of genotype–phenotype relationships that influence biosynthetic pathways. Emerging computational tools—such as Computer-Assisted Structure Elucidation (CASE) systems and integrated multi-omics platforms—further enhance the capacity to model secondary metabolite production and understand how genetic variations affect chemical diversity (SpringerLink, 2025).
In addition to compound discovery, AI is being widely applied in virtual screening and drug design processes, such as predicting molecular docking interactions and binding affinities. These techniques have shown promising results in in silico studies evaluating antioxidant, anti-inflammatory, and antimicrobial activities of phytochemicals (Frontiers in Plant Science, 2023). As these capabilities evolve, AI stands poised to reshape the early stages of drug discovery from plant-based compounds.

Image-Based Identification & Classification

Deep learning approaches, especially convolutional neural networks (CNNs) and Vision Transformers (ViTs), are transforming the identification of plant species and detection of diseases through image analysis. These methods excel at capturing subtle features such as leaf morphology, venation patterns, and texture variations that are often difficult for traditional algorithms or human experts to quantify. The resulting models enable automated, high-throughput phenotyping, crucial for large-scale biodiversity studies and agricultural monitoring (ArXiv, 2024; SpringerLink, 2024).
One landmark study by Mohanty et al. (2016) demonstrated CNN-based crop disease classification with an impressive 99.35% accuracy under controlled laboratory conditions. However, when applied to images collected in real-world field settings—where lighting, background, and disease severity vary—the accuracy dropped significantly to approximately 31%, highlighting challenges related to domain adaptation and data variability (ArXiv, 2016).
Several systematic reviews corroborate the robustness of AI-driven visual recognition systems, emphasizing their ability to accurately discriminate plant diseases across diverse taxa by analyzing visual traits such as leaf texture and venation. These studies underscore AI's promise in revolutionizing plant pathology diagnostics and biodiversity assessments, while also pointing to the need for improved training datasets and adaptable models to enhance real-world performance (ScienceDirect, 2023; Journals.LWW, 2023).

Database-Driven Computational Discovery

Publicly accessible repositories such as Dr. Duke's Phytochemical and Ethnobotanical Database serve as invaluable resources by providing structured, curated information on plant-derived compounds and their associated bioactivities. These databases compile phytochemical profiles, traditional uses, and experimental data, making them ideal for integration into AI and machine learning pipelines (ArXiv, 2023; SpringerLink, 2024). Similarly, digital knowledge libraries—such as India's Traditional Knowledge Digital Library (TKDL)—play a crucial role in preserving centuries-old medicinal plant knowledge in standardized formats, safeguarding cultural heritage while creating rich datasets for computational analysis (Frontiers in Plant Science, 2023; Wikipedia, 2024).
By leveraging these comprehensive data sources, AI models can be trained to predict therapeutic potential and conduct virtual screening of phytochemicals at unprecedented scale and speed. This enables researchers to prioritize promising candidates for experimental validation more effectively, accelerating drug discovery from botanical origins. Moreover, the continual enrichment of these repositories ensures that AI-driven predictions remain grounded in robust, multidisciplinary evidence, facilitating the synthesis of traditional wisdom and cutting-edge computational techniques.

Agricultural & Environmental Integration

Artificial intelligence (AI) is playing an increasingly vital role in plant health monitoring, disease diagnostics, and predictive agriculture. Field imaging powered by AI enables rapid detection of crop diseases and pest infestations through automated analysis of canopy and leaf imagery, facilitating timely interventions. Beyond visual diagnostics, AI-driven microbiome studies unravel complex interactions between plants and their associated microbial communities, offering new avenues for disease resistance and plant health optimization. Additionally, AI models support disease forecasting by integrating environmental and epidemiological data to predict outbreak risks, improving crop management strategies.

Cutting-edge applications include the development of digital twins—virtual models of crops that simulate growth, development, and metabolite production. For instance, Michigan State University is pioneering generative AI-based digital twins of apple trees, enabling researchers to predict phenotypic outcomes under varying environmental conditions (Axios, 2025).

Leading agricultural biotechnology companies are also harnessing AI for trait prediction and drug discovery. Evogene utilizes AI-enabled predictive biology to optimize traits such as yield, stress tolerance, and disease resistance in crops, while Insilico Medicine applies similar AI methodologies to identify drug-like molecules derived from plant metabolic pathways, bridging plant science and pharmaceutical innovation (Wikipedia, 2025).

Challenges & Limitations

Despite the transformative potential of AI in botanical and phytochemical research, several critical challenges must be addressed to fully realize its benefits. First, **data quality and bias** remain fundamental concerns. AI models rely heavily on large, accurate, and well-curated datasets; however, inconsistencies, missing data, or biased sampling can lead to skewed predictions and limit model reliability (ArXiv, 2024). Ensuring high-quality, representative datasets that capture the diversity of plant species and environmental conditions is essential for robust AI applications.

Second, the **generalization** of AI models trained under controlled laboratory or greenhouse settings to real-world field environments remains problematic. Models that perform well in controlled conditions often suffer reduced accuracy when confronted with the variability and noise inherent in natural ecosystems, underscoring the need for diverse, context-rich training data.

Third, **interpretability** and explainability of AI models are critical, especially in pharmacological and clinical contexts where understanding the basis of predictions influences trust and decision-making. The development of explainable AI techniques that elucidate model reasoning without sacrificing performance is an active area of research (Frontiers in Plant Science, 2023).

Finally, **cross-disciplinary integration** poses a significant challenge. Effectively combining computational outputs with ethnobotanical knowledge, phytochemistry, biology, and agronomy demands close collaboration among diverse experts and harmonization of heterogeneous data types. Overcoming these barriers is pivotal for the next generation of AI-driven botanical discovery.

Future Directions

To overcome existing challenges and further enhance AI-driven botanical discovery, several promising approaches and technologies are emerging. Explainable AI (XAI) tools such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) offer ways to demystify complex model predictions by highlighting the contribution of individual features. This transparency helps align AI outputs with known pharmacological markers, increasing trust and facilitating validation by domain experts (Frontiers in Plant Science, 2023).

In data-scarce or partially labeled contexts—common in botanical research—active and semi-supervised learning techniques enable models to iteratively improve by selectively querying the most informative samples, reducing the need for exhaustive annotation.

The future of AI in this field lies in multi-modal frameworks that synthesize heterogeneous data types, including images, omics datasets, chemical fingerprints, and textual ethnobotanical knowledge. Such integrative approaches enable holistic candidate assessment, capturing complex biological and ecological relationships.

Practical field-ready deployment is advancing through smartphone applications like PlantNet, which empower citizen scientists and researchers to perform real-time species and disease identification using models trained on diverse, globally sourced datasets (Wikipedia, 2025).

Finally, collaborative data-sharing initiatives and citizen-science contributions are critical for expanding datasets, improving model generalization, and fostering interdisciplinary synergy. Together, these innovations promise to accelerate and democratize AI-powered botanical exploration.

CONCLUSION

AI-powered discovery is revolutionizing the field of botanical therapeutics by enabling rapid and efficient identification of plant species with high therapeutic potential. By leveraging advanced machine learning algorithms, AI can model complex phytochemical interactions, predict biological activity, and automate species and disease recognition through analysis of diverse data types such as images, omics datasets, and ethnobotanical records. This integrated, data-driven approach is accelerating the pace of natural product discovery far beyond traditional methods.

However, the success and impact of AI in this domain hinge on several critical factors, including the robustness and representativeness of datasets, effective cross-disciplinary collaboration among botanists, chemists, and data scientists, and rigorous experimental validation of computational predictions. Furthermore, responsible deployment practices that address interpretability, bias, and ethical considerations are essential to building trust and ensuring meaningful outcomes.

With ongoing advances in explainable AI and multimodal learning, the future holds great promise. AI is poised to unlock the vast, largely untapped medicinal potential embedded within the botanical world, paving the way for novel therapeutics and sustainable healthcare solutions.

Acknowledgements & Declarations

This is a synthetic review based on compiled knowledge and representative studies.

REFERENCES

1. Abade, A. S., Ferreira, P. A., & Vidal, F. B. (2021). Plant diseases recognition on images using convolutional neural networks: A systematic review. Computers and Electronics in Agriculture, 185, 106125.
2. Alizadehsani, R., Oyelere, S. S., Hussain, S., Calixto, R. R., Albuquerque, V. H. C., Roshanzamir, M., Rahouti, M., & Jagatheesaperumal, S. (2023). Explainable artificial intelligence for drug discovery and development: a comprehensive survey.
3. Axios. (2025). Digital twins in agriculture: AI modeling of apple trees at Michigan State University. Retrieved June 26, 2025.
4. Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311–318.
5. Frontiers in Plant Science. (2023). AI applications in plant disease diagnostics and phytochemical research. Frontiers in Plant Science. Gurulingappa, H., Rajesh Bhat, K., & Ravindran, M. (2022). Machine learning enhances prediction of plants as potential sources of antimalarials. BMC Complementary Medicine and Therapies.
6. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, Article 1419.
7. Pullagura, J. R., Rao, A. N., Kumar, P. S., Raju, M. V., & Madhavi, B. (2025). Grain Quality Analysis Using CNN and Iot Based Strategic Safeguarding. Metallurgical and Materials Engineering, 1273-1278.

8. Sai, R. H. N., Madhavi, B., Reginald, P. J., Kumar, P. S., & Raju, M. V. (2025). Tracing The Pathways and Health Risks of Microplastics in The Food Chain. Metallurgical and Materials Engineering, 1448-1454.
9. Syam Babu, D., Vijay, K., Shakira, S., Mallemkondu, V. S., Barik, P., Kuppam, C., ... & Raju, M. V. (2025). Laterite integrated persulfate based advanced oxidation and biological treatment for textile industrial effluent remediation: optimization and field application. Applied Bionics and Biomechanics, 2025(1), 9325665.
10. MV, R., & K Maria, D. (2020). Green Engineering Strategies on Solid Waste Management to Promote Urban Environmental Sustainability. International Journal of Advanced Science and Technology, 29(5), 4638-4648.
11. Raju, M. V., Palivela, H., Mariadas, K., & Babu, S. R. (2019). Assessment of physico-chemical and biological characteristics and suitability study of lake water: a model study. Technology, 10(01), 1431-1438.
12. MV, R., K Maria, D., & Cyril Lucy, M. (2020). Rapid Environmental Impact Assessment of Lake Water. International Journal of Advanced Science and Technology, 29(3), 8468-8478.
13. Raju, M. V., Mariadas, K., Palivela, H., Ramesh Babu, S., & Raja Krishna Prasad, N. (2018). Mitigation plans to overcome environmental issues: A model study. International Journal of Civil Engineering and Technology, 9(10), 86-94.
14. Satish Kumar, M., Raju, M. V., & Palivela, H. (2017). Comprehensive index of groundwater prospects by using standard protocols-A model study. International Journal of Civil Engineering and Technology (IJCIET), 8(5), 521-526.
15. Hepsibah, P., MV, R., & K Maria, D. (2020). Quality Characteristics of Ground Waters in Few Sources of Industrial Zone. International Journal of Innovative Technology and Exploring Engineering, 9(3), 1134-1137.
16. Kumar, M. S., Raju, M. V., Palivela, H., & Venu Ratna Kumari, G. (2017). Water quality scenario of urban polluted lakes-A model study. International Journal of Civil Engineering and Technology, 8(5), 297-302.
17. Kumar, M. S., Raju, M. V., & Palivela, H. (2017). An overview of managing municipal Solid waste in urban areas-A model study. International Journal of Civil Engineering and Technology, 8(5).
18. Satish Kumar, M., Raju, M. V., Ramesh Babu, S., & Siva Jagadeesh Kumar, M. (2017). Interpretation and correlative study of water simulation in surface water bodies. International Journal of Civil Engineering and Technology (IJCIET), 8(5), 1206-1211.
19. Pullagura, J. R., Osman, O., Kumar, M. S., & Raju, M. V. (2024). LoRaWAN based IoT Architecture for Environmental Monitoring in Museums. Journal of Computational Analysis & Applications, 33(5).
20. Mariadas, K., Rao, T. S., Raju, M. V., & Kumar, M. S. (2024). A Study On the Fluoride Content of the Groundwater in The Gurazala Division of the Palnadu District, Andhra Pradesh, India. Journal of Advanced Zoology, 45(3).
21. Geetha, R., Raju, M. V., Bhavanee, G. P., & Selvakumar, P. (2022). A Review on Medicinal Activity and Its Health Impact of Azadirachta Indica (Neem). YMER., vol. 21(12), 188-200.
22. Reginald, P. J., Raju, M. V., Satyanarayana, D., Mariadas, K., & Kumar, M. S. (2021). Kisaan Seva-Aadhaar Linked Smart Farming Application. Annals of the Romanian Society for Cell Biology, 25(5), 5168-5176.
23. Richard Bollans, A., Aitken, C., Antonelli, A., Bitencourt, C., Goyder, D., Lucas, E., Ondo, I., Pérez Escobar, O. A., Pironon, S., Richardson, J. E., Russell, D., Silvestro, D., Wright, C. W., & Howes, M. J. R. (2023). Machine learning enhances prediction of plants as potential sources of antimalarials. Frontiers in Plant Science, 14, Article 1173328.
24. Richard-Bollans, L., et al. (2023). Machine learning classifiers predict antiplasmodial potential in Apocynaceae, Loganiaceae, and Rubiaceae with higher precision than ethnobotanical methods. Frontiers in Pharmacology.

25. Varghese, J., et al. (2025). Multi-omics integration and AI for metabolite discovery.
26. Varghese, R., Shringi, H., Ramamoorthy, S., & Efferth, T. (2025). Artificial intelligence driven approaches in phytochemical research: trends and prospects. Photochemistry Reviews.
27. Vutukuru, S. S., Asadi, S. S., Vasantha, R. B. V. T., & Raju, M. V. (2012). Plankton biodiversity as indicators of the ecological status of River Moosi, Hyderabad, India. International Journal of Earth Science and Engineering, 5(3), 587-592.
28. Wikipedia contributors. (2024). Plant Net. In Wikipedia, The Free Encyclopedia. Retrieved June 26, 2025.
29. Wikipedia contributors. (2025). Evogene and Insilico Medicine. In Wikipedia, The Free Encyclopedia.
30. Wikipedia contributors. (n.d.). Artificial intelligence in pharmacy. In Wikipedia. Retrieved June 2025.
31. Wikipedia contributors. (n.d.). Plant Net. In Wikipedia. Retrieved June 2025.
32. Zhao, L., Walkowiak, S., & Fernando, W. G. D. (2023). Artificial intelligence: a promising tool in exploring the phytomicrobiome in managing disease and promoting plant health. Plants, 12(9), Article 1852.