Global Research journal of Natural Science and Technology (Grjnst)
Volume: 03 - Issue 4 (2025)
p.ISSN: 2790-7643e.ISSN: 2790-7651

Artificial Intelligence in Phytochemical Screening: Bridging Traditional Knowledge and Modern Drug Discovery – A Review

Kiran Shaheen
BS Biotechnology Student, Department of Bioinformatics and Biosciences, Capital University of Science and Technology, Islamabad, Pakistan
BBT221019@cust.pk , https://orcid.org/0009-0007-5088-1941

Muhammad Aman
BS Biotechnology Student, Department of Bioinformatics and Biosciences, Capital University of Science and Technology, Islamabad, Pakistan
BBT213003@cust.pk , https://orcid.org/0009-0007-2287-1244

Muhammad Asad Anwar (Corresponding Author)
Assistant Professor, Department of Bioinformatics and Biosciences, Capital University of Science and Technology, Islamabad, Pakistan
asad.anwar@cust.edu.pk , https://orcid.org/0000-0002-5915-8940

Abstract

Phytochemical compounds are natural products that have been vital in the discovery of therapeutic agents, narrowing the divide between conventional medicine and drug development, which is hindered by prohibitive expenses, workforce requirements, and low throughput, limiting traditional approaches to phytochemical screening. Recent developments in computational techniques have presented new opportunities to discover bioactive molecules with greater efficiency, providing predictive information on their pharmacological potential. Combining ethnobotanical knowledge, phytochemical databases, and computational tools, researchers can now rapidly identify and optimize plant-derived compounds. In addition to

highlighting this function in crossing traditional medicine with modern remedies, this review emphasizes the latest development in computational approaches for phytochemical screening and addresses difficulties, obstacles and possibilities for the developing field quickly.

Keywords: Phytochemicals, Traditional medicine, Drug discovery, Computational methods, Bioactive compounds**.**

Introduction

Phytochemicals are chemical compounds of biological origin that are produced by plants, and many of them form the basis of the current pharmaceuticals (Tang et al.,2025) (Yang et al.,2024).The science of bioactive compounds Traditionally, the science of bioactive compounds has been thoroughly entrenched in traditional medicine with indigenous knowledge playing a crucial role in the discovery of plants with medicinal value (Mihaylova et al., 2024). Other prominent ones are artemisinin in Artemisia annua, aspirin in willow bark, and paclitaxelin in Taxus species (Bulbul et al., 2025). Although there were these achievements, traditional phytochemical screening systems face drawbacks, such as, complexity of natural product structure, lengthy extraction procedures, and a challenge in large-scale assessment of activities (Feng et al., 2023). The integration of classical knowledge into modern treatment faces significant challenges due to constraints, leading to the development of revolutionary computational techniques for natural product research (Li et al., 2023). These are used to perform rapid structural classification, determine pharmacokinetic properties, and to predict chemical activity. Moreover, the increasing presence of phytochemical databases enhances the capability to integrate traditional medicinal knowledge with new drug discovery efforts (Richard et al., 2023).

Traditional Knowledge and Phytochemical Screening

Traditional medicine has been relied on to prove plant-derived compounds with therapeutic potential. Ethnobotanic's offers useful information on medicinal plants and can inform contemporary research in pharmacology (Parvin et al., 2023) (Bulbul et al., 2025). Some of the most common everyday natural compounds that people may use, including artemisinin (which is extracted from Artemisia annua) as a malaria cure, aspirin (made from willow bark), and paclitaxel (found in Taxus species), can all be traced back to traditional usage (Parvin et al., 2024). Identifying phytochemicals in plants used locally in traditional remedies has been regarded as a practical approach, as any plant with bioactivity has been previously linked to it (Parvin et al., 2025). The repository of knowledge across generations will provide a starting framework for selection, making it less expensive and time-consuming to screen a large number of people (Al-Sammarraie et al., 2025). Nonetheless, conventional phytochemical-screening poses a considerable challenge. Many extraction techniques differ in their efficiency, leading to variations in the yield and quality of the compound obtained during extraction (Tang et al., 2025). Traditional bioassay-directed isolation methods are both economically and logistically costly, and they cannot meet the growing need for new therapeutics (Mirakhori et al., 2025). Also, numerous phytochemicals are present in minute quantities in plant matrices, making them difficult to isolate and structure-elucidate (Gaudenico et al., 2023). Recent developments have enhanced the standardization of conventional screening approaches. Recent methods of extraction include: ultrasound-assisted extraction, microwave-assisted extraction, and supercritical fluid extraction, where modern methods are improved to enhance yields and stability of the compound (Tang et al., 2025). These approaches bridge the gap between ancient

knowledge and modern pharmacological needs, accelerating the discovery of bioactive natural products (Shoaib et al., 2023).

Computational Approaches in Phytochemical Screening

Phytochemical studies are actively starting to use machine learning (ML) techniques to process large datasets, predicting compound behavior and classifying molecular geometries (Avellaneda et al., 2025). Various algorithms like Support Vector Machines (SVM), decision tree, and random forest have been found relevant to QSAR modeling and consequently have enabled scientists to forecast biological activity, toxicity, and drug-likeness of phytochemicals (Ghosh et al., 2025). Recent works revealed that ML is both effective to predict potential antimalarial plant species, with higher predictive accuracy than traditional experimental screening (Richard et al., 2023). Moreover, dimensionality reduction methods and clustering algorithms support large-scale chemical library organisation, allowing structural and functional similarities of phytochemicals to be identified (Murmu et al., 2024). By prioritizing promising molecules for additional experimental validation, these computational methods lower workload and expenses. Deep learning (DL) algorithms are ML methods with expanded functionality to show highly non-linear associations in chemical and biological data sets (Rustandi et al., 2023). Models: Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNNs) can be generally parallelized to accurately analyze the structural data of natural compounds and predict their pharmacological actions (Arji et al., 2019). DL approaches are also integrated into plant bioinformatics workflows, including image-based recognition of plant phenotypes, metabolite profiling, and prediction of phytochemical biosynthetic pathways (Basnet et al., 2025).

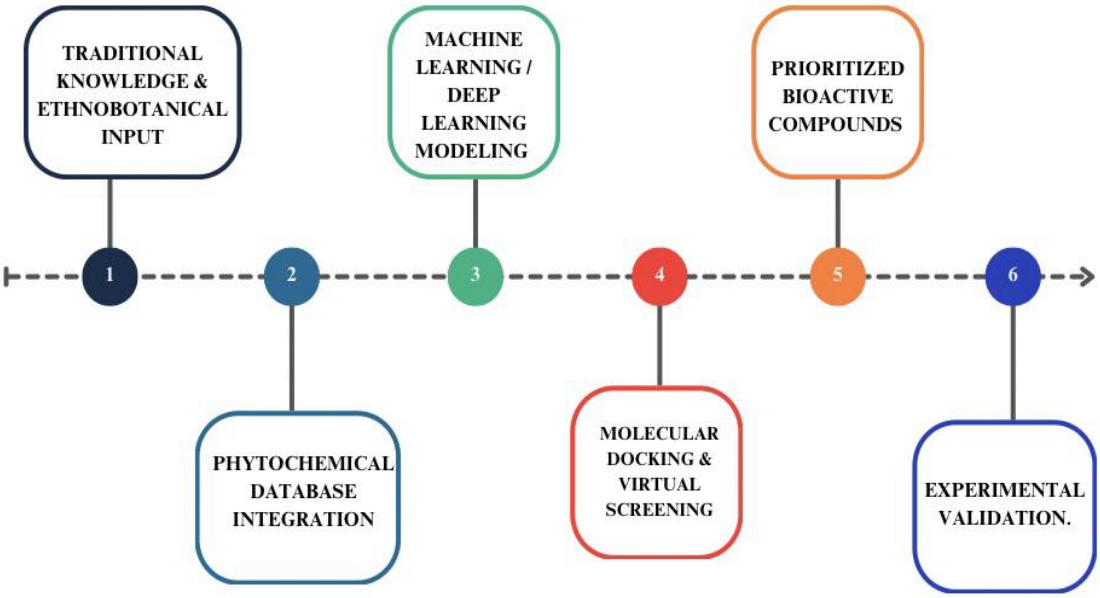

Figure1: Workflow of Computational Phytochemical Screening (Basnet et al., 2025).

In the traditional black-box nature of deep learning models, explainable DL models are being developed to offer better transparency and interpretability than predictive modeling (Basnet et al., 2025). Simplifying the computational screening method has been made possible by the growth of curated phytochemical databases. Large compendia of annotated phytochemicals are available in NPASS, IMPPAT, Super-Natural II, and COCONUT, which are used to submit datasets using other resource singers (Mullowney et al., 2023) (Pandey et al., 2025). To forecast the binding affinities of phytochemicals and target proteins, these databases are usedin virtual screening procedures and computer docking platforms (Dhudum et al.,

2024). Additionally, the application of natural language processing (NLP) and text mining has been used to determine pertinent phytochemical knowledge in ethnopharmacological literature, which has largely reinforced the incorporation of traditional medicinal knowledge alongside contemporary drug discovery Comprehensively, the entire field of phytochemical research has been revolutionized-through the use of computational tools, which provide more rapid and precise CSBA predictions/optimization, less expensive experiments, and the potential to fast-track the discovery process. Combined with cheminformatics, molecular docking, and curated databases (e.g., PubChem and ChEMBL), ML and DL models improve the use of virtual screening as well as prioritization of compounds in drug discovery (Tang et al., 2025). These advances aside, limitations in data coverage, standardization, and interpretability suggest that the creation of more streamlined and comprehensive phytochemical resources is still required (Parvin et al., 2025).

Table 1. Key Machine Learning and Deep Learning Methods in Phytochemical Screening

Methods	Purpose	Citations
Support Vector Machines (SVM)	Predict biological activity and toxicity	(Ghosh et al., 2025)
Decision Trees & Random Forests	Classification and regression tasks	(Ghosh et al., 2025)
Convolutional Neural Networks (CNN)	Structural analysis and image-based phenotyping	(Arji et al., 2019)
Graph Neural Networks (GNN)	Capture structural relationships in chemical data	(Basnet et al., 2025)
Natural Language Processing (NLP)	Extract knowledge from literature	(Rustandi et al., 2023)

Challenges and Limitations

Despite progress, several barriers restrict the widespread use of computational methods in phytochemical research. A major constraint is the scarcity of large, high-quality phytochemical datasets. Data inconsistencies and incomplete annotations reduce model training efficiency and accuracy (Feng et al., 2023). Natural products exhibit high structural diversity and stereoisomerism, complicating computational modeling and similarity-based predictions (Bulbul et al., 2025). Models trained on limited datasets may not generalize across diverse plant species. Overfitting is a frequent issue, limiting predictive power for novel phytochemicals (Murmu et al., 2024). Computational predictions often lack experimental confirmation.

Variability in experimental conditions and absence of benchmarking standards hinder reproducibility (Mullowney et al., 2023). The incorporation of traditional knowledge into computational platforms raises ethical concerns about intellectual property rights and benefit-sharing with indigenous communities (Basnet et al., 2025).

Future Perspectives

Future phytochemical screening will increasingly integrate multi-omics data, including genomics, transcriptomics, proteomics, and metabolomics (Gaudencio et al., 2023). This integration allows mapping of biosynthetic pathways, prediction of active metabolites, and linking chemical profiles to therapeutic effects, accelerating novel compound discovery, and enhancing understanding of plant metabolism. Despite high predictive accuracy, model interpretability remains a challenge. Explainable computational frameworks can clarify why a phytochemical is predicted to be active, improving trust and supporting experimental validation (Parvin et al., 2024). Expanding and standardizing phytochemical databases, including chemical structures, biological activities, ethnobotanical knowledge, and pharmacokinetic data, is essential for robust model training (Dhudum et al., 2024). Future efforts should also address ethical considerations, fair use of traditional knowledge, and sustainable sourcing of plant materials to ensure equitable and responsible drug discovery (Basnet et al., 2025).

Conclusion

Phytochemicals remain a vital source of therapeutic agents, bridging the gap between traditional medicine and modern drug discovery. Conventional screening methods, while valuable, are limited by time, cost, and experimental complexity. Computational approaches

have emerged as essential tools for accelerating phytochemical research, enabling the prediction of bioactivity, structural analysis, and efficient prioritization of compounds. Integration of computational models with multi-omics data, comprehensive databases, and predictive frameworks enhances the discovery pipeline while providing deeper insights into plant metabolism and chemical diversity. The future of these approaches is filled with challenges like lack of data, complexity of structure, interpretability of models and ethical concerns that need to be addressed in order to maximize the potential of such methods.

References

Al-Sammarraie, R. N., et al. (2025). An artificial intelligence-aided scoping review of medicinal plant research in the Fertile Crescent. *Frontiers in Pharmacology, 16,* 1542709. https://doi.org/10.3389/fphar.2025.1542709

Avellaneda-Tamayo, J. F., Agudo-Muñoz, N. A., Sánchez-Galen, J. E., López-Pérez, J. L., & Medina-Franco, J. L. (2025). Chemoinformatic characterization of NAPROC-13: A database for natural product 13C NMR dereplication. *Discover Applied Sciences, 7*(1), 491. https://doi.org/10.1007/s42452-025-06772-1

Basnet, B. B., Zhou, Z. Y., Wei, B., & Wang, H. (2025). Advances in AI-based strategies and tools to facilitate natural product and drug development. *Critical Reviews in Biotechnology.* https://doi.org/10.1080/07388551.2025.2478094

Bülbül, E. F., Bode, H. B., Schmitt, S., & Bozhüyük, K. A. J. (2025). Engineering the future of medicine: Natural products, synthetic biology and artificial intelligence for next-generation therapeutics. *Clinical and Translational Medicine, 15*(2), e70146. https://doi.org/10.1002/ctm2.70146

Dhudum, R., Ganeshpurkar, A., & Pawar, A. (2024). Revolutionizing drug discovery: A comprehensive review of AI applications. *Drugs and Drug Candidates, 3*(1), 148–171. https://doi.org/10.3390/ddc3010009

Feng, Y., Cao, X., Li, W., Lin, H., & Jin, J. (2023). Recent advances of diffusion models in image processing. *arXiv*. https://doi.org/10.48550/arXiv.2308.10921

Gaudêncio, S. P., et al. (2023). Advanced methods for natural products discovery: Bioactivity screening, dereplication, metabolomics profiling, genomic sequencing, databases and informatic tools, and structure elucidation. *Marine Drugs, 21*(5), 308. https://doi.org/10.3390/md21050308

Ghosh, S., Basu, S., Kayal, T., Ashok, G., Ramaiah, S., & Anbarasu, A. (2025). Computational advancements to facilitate therapeutic application of phytochemicals: Where do we stand? *Discover Applied Sciences, 7*(1), 491. https://doi.org/10.1007/s42452-025-06772-1

Li, Y., Dang, Y., Li, S., Rodriguez, F., Wang, S., Zhou, Y., He, D., & Chen, R. (2023). Natural products in prevention and treatment of COVID-19: Current progress and future perspectives. *Journal of Medicinal Chemistry, 66*(6), 4312–4344. https://doi.org/10.1021/acs.jmedchem.2c01257

Mihaylova, S., Tsvetkova, A., Georgieva, E., & Vankova, D. (2024). Bioactive phyto-compounds with antimicrobial effects and AI: Results of a desk research study. *Microorganisms, 12*(6), 1055. https://doi.org/10.3390/microorganisms12061055

Mirakhori, F., & Niazi, S. K. (2025). Harnessing the AI/ML in drug and biological products discovery and development: *The regulatory perspective. Pharmaceuticals, 18*(1), 47. https://doi.org/10.3390/ph18010047

Mullowney, M. W., et al. (2023). Artificial intelligence for natural product drug discovery. Nature Reviews *Drug Discovery, 22*(11), 895–916. https://doi.org/10.1038/s41573-023-00774-7

Murmu, S., Sinha, D., Chaurasia, H., Sharma, S., Das, R., Jha, G. K., & Archak, S. (2024). A review of artificial intelligence-assisted omics techniques in plant defense: Current trends and future directions. *Frontiers in Plant Science, 15,* 1292054. https://doi.org/10.3389/fpls.2024.1292054

Parvin, N., & Mandal, T. K. (2024). Computational phytochemistry: Advancing natural product research through artificial intelligence and machine learning. *Processes, 12*(12), 2488. https://doi.org/10.3390/pr122412488

Parvin, N., Aslam, M., Joo, S. W., & Mandal, T. K. (2025). Nano-phytomedicine: Harnessing plant-derived phytochemicals in nano carriers for targeted human health applications. *Molecules, 30*(15), 3177. https://doi.org/10.3390/molecules30153177

Parvin, N., Joo, S. W., & Mandal, T. K. (2023). Artificial intelligence in natural product drug discovery: A systematic review. *PLoS ONE, 18*(5), e0285716. https://doi.org/10.1371/journal.pone.0285716

Pandey, B. S., Durgapal, S., Kumar, P., et al. (2025). Unraveling the artificial intelligence role in drug discovery and pharmaceutical product design: An opportunity and challenges. *Discover Artificial Intelligence, 5,* 72. https://doi.org/10.1007/s44163-025-00330-9

Richard-Bollans, A., et al. (2023). Machine learning enhances prediction of plants as potential sources of antimalarials. *Frontiers in Plant Science, 14,* 1173328. https://doi.org/10.3389/fpls.2023.1173328

Rustandi, T., Prihandiwati, E., Nugroho, F., Hayati, F., Afriani, N., Alfian, R., Aisyah, N., Niah, R., Rahim, A., & As-Shiddiq, H. (2023). Application of artificial intelligence in the development of Jamu 'traditional Indonesian medicine' as a more effective drug. *Frontiers in Artificial Intelligence, 6,* 1274975. https://doi.org/10.3389/frai.2023.1274975

Shoaib, M., Shah, B., Sayed, N., Ali, F., Ullah, R., & Hussain, I. (2023). Deep learning for plant bioinformatics: An explainable gradient-based approach for disease detection. Frontiers in Plant Science, 14, 1283235. https://doi.org/10.3389/fpls.2023.1283235

Tang, Z.-X., Ye, X.-P., Jiang, Z.-B., Chen, Y.-X., Shen, H.-Y., Hu, Y.-Y., & Shi, L.-E. (2025). Recent advances in flavonoids from the genus Polygonatum: Extraction and processing methods. *Frontiers in Sustainable Food Systems, 9.* https://doi.org/10.3389/fsufs.2025.1569034

Yang, J. S., Tsai, S. C., Hsu, Y. M., Bau, D. T., Tsai, C. W., Chang, W. S., Kuo, S. C., Yu, C. C., Chiu, Y. J., & Tsai, F. J. (2024). Integrating natural product research laboratory with artificial intelligence: Advancements and breakthroughs in traditional medicine. *Biomedicine 14*(4), 1–14. https://doi.org/10.37796/2211-8039.1475