RESEARCH ARTICLE

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN MODERN PHARMACOGNOSY: EXPANDING THE FRONTIERS OF NATURAL DRUG DISCOVERY

Sujitha R[1] | Vishnupriya S[1] | Subasri S | Jaya Krishnan V[1] | Rajasekaran A[1] * | | Arivukkarasu R |

[1]KMCH College of Pharmacy, Coimbatore, Tamil Nadu, India-641048 (The Tamil Nadu Dr. M.G.R. Medical University, Chennai, Tamil Nadu, India-600036)

ABSTRACT

Pharmacognosy, the scientific discipline concerned with drugs obtained from natural sources, is using the combination of machine learning (ML) and artificial intelligence (AI) to evolve quickly. These technologies provide advanced computational approaches for managing large and complex datasets related to medicinal plants, phytochemicals, and biological activities. AI-based systems support plant identification, phytochemical profiling, bioactivity prediction, toxicity assessment, and molecular modeling. Machine learning algorithms enable predictive modeling and virtual screening, thereby accelerating natural product–based drug discovery [1, 2]. In addition, AI contributes to quality control, standardization, and formulation development of herbal medicines, ensuring safety, efficacy, and reproducibility [3]. The convergence of pharmacognosy, data science, and pharmaceutical technology represents a major shift toward a more predictive, efficient, and sustainable approach to natural drug research and development.

Keywords: artificial intelligence, machine learning, pharmacognosy, herbal drug development, phytochemical analysis, medicinal plant

***Corresponding Author**

Arivukkarasu R.
KMCH College of Pharmacy,Coimbatore, Tamil Nadu, India-641048.
E-mail: phytoarivu@gmail.com

INTRODUCTION

The scientific field of pharmacognosy, which deals with the drugs of natural origin, is rapidly advancing with the integration of artificial intelligence (AI) and machine learning (ML). These technologies enable efficient processing of large volumes of data related to medicinal plants, phytochemicals, and biological activities. AI-based applications support plant identification, prediction of bioactivity, and molecular modeling, thereby enhancing natural product drug discovery. However, traditional pharmacognostic techniques are often time-consuming, labor-intensive, and limited in their ability to handle complex and large-scale data. This creates a pressing need for advanced computational approaches. Therefore, AI and MI offer effective solutions to improve efficiency, accuracy, and scalability in modern pharmacognosy research [5].

FUNDAMENTALS OF AI AND ML IN PHARMACEUTICAL RESEARCH

The artificial intelligence systems work based on the idea that by training computational models, researchers can learn about patterns and make predictions by using large amounts of data. ML algorithms are divided into supervised learning, unsupervised learning, and deep learning. [4]

Pharmaceutical research: Pharmaceutical ML models have been utilized in: Drug target identification Lead compound optimization Pharmacokinetic and toxicity prediction Structure activity relationship (SAR) analysis [6]

TYPES OF ML[Table .No:1]

ML Type	Description	Pharmacognosy Example
Supervised Learning	Uses labeled data for prediction	Predicting anti-inflammatory activity of plant compounds
Unsupervised Learning	Finds patterns in unlabeled data	Clustering phytochemicals based on structure
Deep Learning	Neural networks for complex data	Image based medicinal plant identification

WORKING FLOW CHART

The process begins with raw data, which may include genomic sequences, chemical libraries, spectroscopic signatures, or plant image datasets. This raw data then undergoes data preprocessing, a critical step that involves cleaning, normalization, handling missing values, and removing noise or irrelevant information to ensure data quality and consistency. Following preprocessing, feature extraction is performed to transform the cleaned data into meaningful, reduced-dimensional representations such as molecular fingerprints, spectral peaks, or morphological traits that capture the most relevant patterns for the task at hand. These extracted features are subsequently fed into the model training phase, where machine learning or quantum-classical hybrid algorithms learn underlying relationships, optimize parameters, and build predictive or generative models. The trained model then proceeds to validation, where its performance is rigorously assessed using independent test sets, cross-validation, or external benchmarks to evaluate accuracy, robustness, and generalizability. Once validated, the model moves to deployment, where it is integrated into real-world platforms, software tools, or laboratory information systems for routine use. Finally, the pipeline culminates in the output stage, delivering actionable results such as accelerated drug discovery (e.g., candidate compounds, binding affinity predictions) or accurate plant identification (e.g., species classification, chemotaxonomic markers), thereby translating raw data into tangible scientific or clinical value.

DATABASES SUPPORTING AI-DRIVEN PHARMACOGNOSY

The AI applications in pharmacognosy rely on databases. The tools and databases such as PubChem, Chambly, TCMsp, and Knapsack are important, and they make chemical structures, biological activities, and

ethnopharmacological information available. These databases provide molecular descriptors, fingerprints, and biological annotations, which can be used to train the ML models [7, 8].

Databases for AI in Pharmacognosy [Table.No:2]

NAME	TYPE	URL	USE IN AI
PUBCHEM	Chemical database	https://pubchem.ncbi.nlm.nih.gov	Molecular data for ML training
CHEMBL	Bioactivity database	https://www.ebi.ac.uk/chembl	Drug target prediction
TCMSP	Traditional Chines Medicine	https://tcmsp-e.com	Herbal compound analysis
KNAPSACK	Metabolite database	http://www.knapsackfamily.com	Plant metabolite relationships

AI Integration in Pharmacognosy

The process begins with specialized databases such as PubChem and TCMSP (Traditional Chinese Medicine Systems Pharmacology Database), which serve as primary sources of chemical compounds, natural product information, and pharmacological data. These databases feed into machine learning (ML) and deep learning (DL) models, where algorithms are trained to recognize complex patterns, predict molecular properties, and generate novel drug-like candidates. To implement and operationalize these models, a suite of computational tools is employed, including Python as the core programming language and RDKit as the cheminformatics library for molecular manipulation, fingerprint generation, and structure analysis. The integration of these ML/DL models with the aforementioned tools enables efficient data processing, feature engineering, and model optimization. Finally, the pipeline converges on the output stage, supporting applications in drug discovery (e.g., lead identification, ADMET prediction, binding affinity estimation) and quality control (QC), such as verifying compound purity, consistency in natural product extracts, or batch-to-batch reproducibility in pharmaceutical manufacturing. Thus, Figure 2 illustrates a cohesive ecosystem where public databases, advanced machine learning techniques, and specialized cheminformatics tools jointly accelerate research and quality assurance outcomes.

SOFTWARE AND COMPUTATIONAL TOOLS

4.1 Programming and Data Analysis

Python is popular in terms of data analysis and development of AI models. NumPy, pandas, SciPy, and scikit-learn are Python libraries that offer a full ecosystem of machine learning and data science [9]. R is especially applicable in statistical computation and in data mining [10].

4.2 Image Analysis

TensorFlow and OpenCV are common in the creation of deep learning-based image analysis models in the classification and authentication of medicinal plants. Moreover, image classification models can also be developed and trained in the simplest possible way using no-code models like Teachable Machine, thereby allowing plant identification with as little knowledge of software development as needed. The tools aid quick and correct identification of plant species based on visual appearance. [11–13].

4.3 Cheminformatics Tools

Molecular descriptors, fingerprints, and structural features necessary in QSAR and machine learning research are often computed using RDKit and PaDEL. RDKit is used to do molecular editing and detect functional groups, whereas PaDEL offers a collection of many chemical descriptors. The two of them are used to enable the efficient representation of chemical structures to be used in predictive modeling. [13–15].

4.4 Molecular Modeling and Docking

The predictions of ligand-receptor interactions and binding affinities of the phytochemicals with the biological targets are made by use of AutoDock, which helps in virtual screening. Machine learning and MATLAB libraries like scikit-learn can help in simulation, data analysis, and model development, therefore improving structure-based natural product drug discovery. [16–18].

METHODOLOGY FOR AI-BASED PHARMACOGNOSY STUDIES

5.1 Data Collection

Data collection entails the use of credible databases and research findings to obtain plant pictures and chemical frameworks and ethnobotanical facts, as well as biological activity data. These sources of information can be used to construct extensive datasets to conduct AI-based pharmacognosy studies.

5.2 Data Preprocessing

Preprocessing of data involves the cleaning of data sets and elimination of errors, inconsistencies, and missing values and data normalization to ensure uniformity. It is also the process of deriving pertinent data out of unprocessed data such that significant data is achieved. These measures enhance the quality of the data and make the information appropriate and useful to the machine learning analysis.

5.3 Model Training

The training process of models can be applied by feeding the machine learning algorithms with labeled datasets to enable the models to learn trends, relationships, and patterns in the data. Through this process, predictive models are developed and able to analyze the pharmacogenetic data.

5.4 Model Evaluation

Evaluating the performance of trained machine learning and deep learning models based on standard measures of performance such as accuracy, precision, recall, and validation sets. This is done to make sure that the models established to identify medicinal plants and also in other applications of pharmacogenetics give reliable and reproducible results.

5.5 Deployment

Deployment is the process of implementing trained and verified machine learning models into a real platform, whether it is a desktop program or mobile app or even a laboratory system. This will enable real-time use of models to identify medicinal plants and other pharmacogenetic analyses. [19, 20].

CASE STUDY

Title: AI predicts antidiabetic compounds from

- Ficus religiosa
- Researchers are looking for natural products to cure diabetes.
- Rather than just doing lab work (which takes years), they used artificial intelligence (AI) to discover the plant chemical.
- Ficus religiosa (Peepal tree) is used in traditional medicine. Known to have anti-diabetic effects
- It contains:
- Flavonoid, Alkaloid, Phenol, Steroid. These chemicals may help control glucose.
- They focused on an enzyme: DPP-4 (dipeptidyl peptidase-4).
- DPP-4 breaks down hormones that stimulate insulin. If you block DPP-4, more insulin is released, and blood sugar decreases.

AI contribution

Collected plant compounds (from databases)

Predicted drug properties (ADMET): absorption, toxicity, safety

Used deep learning (DeepBindGCN) to estimate the binding affinity to DPP-4d molecular docking. Models the compound's interaction with the enzyme

Energy calculation: Checks how stable the binding

So basically, AI = rapid screening + prediction [27].

STUDIES TABLE [Table.No:3]

PLANT	AI METHOD	OUTCOME	YEAR
CURCUMA LONGA	Random Forest	Anti diabetic prediction	2022
AZADIRACHTA INDICA	SVM	Antimicrobial activity	2021
OCIMUM SANCTUM	Deep learning	Compound classification	2023
WITHANIA SOMNIFERA	ANN	Anti cancer prediction	2024
ALOE VERA	CNN	Plant identification	2022
FICUS RELIGIOSA	Deep learning	Anti diabetic prediction	2023

COST EFFECTIVENESS SECTION

The incorporation of AI in drug discovery and development substantially cuts the cost associated with conventional drug discovery by reducing the number of experiments, time taken to identify leads, and resource allocation. Automated data processing and virtual screening remove the need for multiple wet-lab experiments, leading to cost-effectiveness of AI in pharmacognosy and drug development.

FEDERATED LEARNING

Federated learning allows various organizations to jointly develop machine learning models while keeping their data private. This is a useful approach in pharmacognosy where there are issues with data privacy, ownership, and

geographical distribution of medicinal plant data. It enables the development of powerful global models without compromising data security.

APPLICATIONS IN PHARMACOGNOSY

9.1 Medicinal Plant Identification

The deep learning models are also being applied in the identification of medicinal plants through the analysis of leaf, root, and flower images. Such models acquire the predictions of complex visual patterns and visual features to allow proper recognition of species and facilitate the discovery of medicinal applications of plant-derived natural products. These methods enhance the speed and accuracy of plant authentication in pharmacognosy studies. [21].

9.2 Phytochemical Prediction

Machine learning methods are used to identify bioactive classes of compounds as well as possible pharmacological action using chemical and biological data. Phytochemical studies that are supported by AI are beneficial in the interpretation of the relationship between compounds and their functions and as a strategy to determine the potential of promising phytochemicals to be further investigated.[23].

9.3 Drug Lead Discovery

It is achieved by AI-based virtual screening and predictive modeling, so the identification of potential lead compounds in large natural product libraries is speeded up. These methods also minimize the labor involved in the experiment and improve efficiency in the initial discovery of drugs based on plant materials.[23].

9.4 Quality Control and Standardization

Artificial intelligence is relevant in the authentication of herbal drugs, detection of adulterants, and determination of quality standards. The AI-based analytical tools enhance the consistency, safety, and reliability of herbal medicines and aid in regulatory compliance and standardization. [22].

9.5 Metabolomics and Pathway Prediction.

The method of predicting biosynthetic pathways of secondary metabolites and the analysis of metabolomics data are done using deep learning methodologies. These approaches help to study multifaceted metabolic networks and help to identify new bioactive molecules and biosynthetic pathways. [24, 25].

Deep Learning in Metabolomics

Deep learning, a type of artificial intelligence, employs multi-layered neural networks that learn representations from data. Within metabolomics, it's used to process and interpret the high-dimensional data obtained from mass spectrometry (MS) and nuclear magnetic resonance (NMR). [28].

Application

- Metabolite identification, biomarker discovery, classification of plant species, and pathway analysis in medicinal plants
- Role in Metabolomics Deep learning enables:
- Automated extraction of features from spectral data, identification of unknown metabolites, Detection of hidden patterns and relationships Classification of metabolomic profiles.
- Typical models include convolutional neural networks (CNNs) for spectral/image data or autoencoders for dimensionality reduction and denoising. [29]

PHARMACY-ORIENTED PERSPECTIVES

10.1 Role in Herbal Drug Development

Artificial intelligence supports herbal drug development by assisting in the selection of suitable plant sources, optimization of extraction processes, and rational design of formulations. AI-based analysis helps identify promising bioactive compounds and improves decision-making throughout the drug development process.

- Explainable AI (XAI) in herbal medicine
- Explainable AI is used to explain AI predictions rather than a "black box." For herbal medicine, where herbs have many active compounds that work on different targets, XAI allows researchers to
- Discover the active compounds that work.
- Understand how herbs work (mechanism of action)
- Analyze herbal combinations and formulations.
- Support safe and personalized treatment
- XAI in general enhances transparency, credibility, and scientific evidence of herbal medicines and bridges traditional medicine with modern science. [30]
- The Role of Explainable AI (XAI) in Herbal Medicine - In Brief Explainable AI (XAI) is a way to provide explanations of AI models rather than "black boxes."
- This is crucial in herbal medicine because herbs have many active compounds and multiple targets and pathways [31].

10.2 Integration with Pharmaceutics

AI is increasingly integrated with pharmaceutics to predict excipient compatibility, dissolution behavior, and stability of herbal formulations. These predictions help in developing safe, effective, and stable dosage forms while reducing trial-and-error experimentation.

10.3 Clinical Pharmacy Relevance

Machine learning tools help identify potential herb–drug interactions and support personalized herbal therapy. These applications enhance patient safety and assist pharmacists in selecting appropriate herbal treatments based on individual needs.

10.4 Regulatory and quality assurance

AI-generated analytical fingerprints and predictive models assist in meeting pharmacopeial and Good Manufacturing Practice (GMP) requirements. This supports regulatory compliance and ensures consistent quality of herbal medicinal products.

CHALLENGES AND LIMITATIONS

Despite its advantages, the application of artificial intelligence in pharmacognosy faces several challenges. These include limited availability of high-quality and well-annotated datasets, lack of standardization across data sources, and the black-box nature of some machine learning models, which makes interpretation difficult. In addition, successful implementation requires interdisciplinary expertise combining pharmacognosy, data science, and computational biology.

FUTURE PROSPECTS

Future developments in AI-driven pharmacognosy are expected to include multi-omics data integration, development of explainable AI models, establishment of global herbal databases, and implementation of smart cultivation systems. Artificial intelligence will play an important role in promoting sustainable cultivation of

medicinal plants by optimizing growth conditions, monitoring plant health, and improving yield and quality. These advancements will strengthen conservation efforts and support sustainable and efficient natural drug development [26].

CONCLUSION

Pharmacognosy is benefiting from the application of artificial intelligence and machine learning, which support data-driven, accelerated natural drug discovery. Their use improves precision, speeds up the process, and enables sustainable use of nature's resources. Emerging developments in explainable AI and data-sharing platforms will further support their integration in evidence-based herbal drugs.

REFERENCES

1. Kinghorn AD. Pharmacognosy in the 21st century. J Pharm Pharmacol. 2001;53:135-148.
2. Khanzode KU, Sarode RD. Advantages and disadvantages of artificial intelligence and machine learning: A literature review. Int J Libr Inf Sci. 2020;9(1): 3.
3. Sinoriya P, Sharma R, Patel M. The role of artificial intelligence in modern pharmacognosy: advances and applications. J Appl Pharm Sci Res. 2024;7(3):10-21.
4. Zhou ZH. Machine learning. Cham: Springer Nature; 2021.
5. Baishyaa D, Sharma P, Das S. Molecular pharmacognosy in the 21st century: a multidisciplinary approach for natural drug discovery. J Nat Prod Res. 2023;37(5):621-635.
6. Muthuswamy R, Mozhi M, Kamaleeshwari R, Kowshika G. The role of artificial intelligence, machine learning, and bioinformatics in revolutionizing drug development. Kronika J. 2025;25(6):1-12.
7. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. 2009;37(Web Server issue):W623-W633. doi:10.1093/nar/gkp456.
8. Kim S. Exploring chemical information in PubChem. Curr Protoc. 2021;1(8):e217. doi:10.1002/cpz1.217.
9. Navlani A, Fandango A, Idris I. Python data analysis. Birmingham: Packt Publishing; 2021.
10. Wong JJN, Nik F. Development of species recognition models using Google Teachable Machine on shorebirds and waterbirds. J Taibah Univ Sci. 2022;16(1):1096-1111. doi: 10.1080/16583655.2022.2105055.
11. Budhathoki A, Shrestha A. Image classification through no-code Teachable Machine platform. Kathmandu: ResearchGate Preprint; 2022.
12. Lamsal RR, Lamsal S. Introducing image classification through no-code Teachable Machine platform: an intuitive approach for novice learners. 2025.
13. Bjerrum EJ, Palunas K, Menke J. Python-based interactive RDKit molecule editing with rdEditor. J Cheminform. 2024;16(1):1-9.
14. Colmenarejo G. EFGs: a complete and accurate implementation of Ertl's functional group detection algorithm in RDKit. J Chem Inf Model. 2025;65(3):1061-1066.
15. Bento AP, Gaulton A, Hersey A, et al. An open-source chemical structure curation pipeline using RDKit. J Cheminform. 2020;12:51. doi:10.1186/s13321-020-00436-5.
16. Pacheco AB. Introduction to AutoDock and AutoDock tools. Baton Rouge (LA): Louisiana State University; 2012.
17. Gilat A. MATLAB: an introduction with applications. Hoboken (NJ): John Wiley & Sons; 2017.
18. Garreta R, Moncecchi G. Learning scikit-learn: machine learning in Python. Birmingham: Packt Publishing; 2013.
19. Jolly K. Machine learning with scikit-learn quick start guide. Birmingham: Packt Publishing; 2018.
20. Rao RU, Kumar A, Reddy M. Identification of medicinal plants using deep learning. Int J Res Appl Sci Eng Technol. 2022;10(4):306-322.
21. Yoo S, et al. A deep learning-based approach for identifying the medicinal uses of plant-derived natural compounds. Front Pharmacol. 2020;11:584875. doi:10.3389/fphar.2020.584875.

22. Kaur R, Baldi A, Baldi A. Artificial intelligence in herbal drug authentication: revolutionizing identification, adulterant detection, and standardization. Curr Comput Sci. 2025;4(1):E29503779378012.
23. Varghese R, et al. Artificial intelligence-driven approaches in phytochemical research: trends and prospects. Phytochem Rev. 2025;24(1):1-16.
24. Bao H, et al. Prediction of plant secondary metabolic pathways using deep transfer learning. BMC Bioinformatics. 2023;24:348. doi:10.1186/s12859-023-05464-z.
25. Lee WY, et al. A machine learning approach for recommending herbal formulae with enhanced interpretability and applicability. Biomolecules. 2022;12(11):160 4. doi:10.3390/biom12111604.
26. Mahato R, et al. Application of artificial intelligence in sustainable cultivation of medicinal plants. Environ Sustain. 2023;6(4):773-784.
27. Alam, Md Ashad, and Md Amanullah. "Artificial Intelligence Powered Identification of Potential Antidiabetic Compounds in Ficus religiosa." *arXiv preprint arXiv:2510.19867* (2025).
28. Sen, Partho, et al. "Deep learning meets metabolomics: a methodological perspective." *Briefings in Bioinformatics* 22.2 (2021): 1531-1542.
29. Pomyen, Yotsawat, et al. "Deep metabolome: Applications of deep learning in metabolomics." *Computational and Structural Biotechnology Journal* 18 (2020): 2818-2825.
30. Ding, Qingyao, et al. "Explainable artificial intelligence in the field of drug research." *Drug Design, Development and Therapy* (2025): 4501-4516.
31. Fatima, Syeda Urfia, and S. Shailaja. "AI in Traditional and Herbal Medicine." *Journal of Scientific Research and Technology* (2025): 114-121.