"Artificial Intelligence in Pharmacognosy: Revolutionizing Medicinal Plant Identification, Phytochemical Discovery and Herbal Drug Standardization"

Nikhil D. Ghorpade[1], Aadi Sathyan[2], Diksha C. Mhatre[3], Ashutosh P. Kamble[4], Suchita S. Shete[5], Rashmi Sharma[6]

[1]*Asst. Professor, Department of Pharmacology, SGVSS Sakeshwar College of Pharmacy Chas,*
[2]*Asso. Professor, Department of Pharmacy, Nirmala College of Health Science Chalakudy,*
[3]*Asst. Professor, Department of Pharmacology, Bombay Institute of Pharmacy and Research,*
[4]*Asst. Professor, Department of Pharmacology, D.Y. Patil college of Pharmacy, Kadam wadi, Kolhapur,*
[5]*Asst. Professor, Department of Pharmacology, Jaihind College of Pharmacy,*
[6]*Asst. Professor, Department of Biochemistry, Shri Vaishnav Vidhyapeeth Vishwavidyalaya, Indore.*

Graphical Abstract—

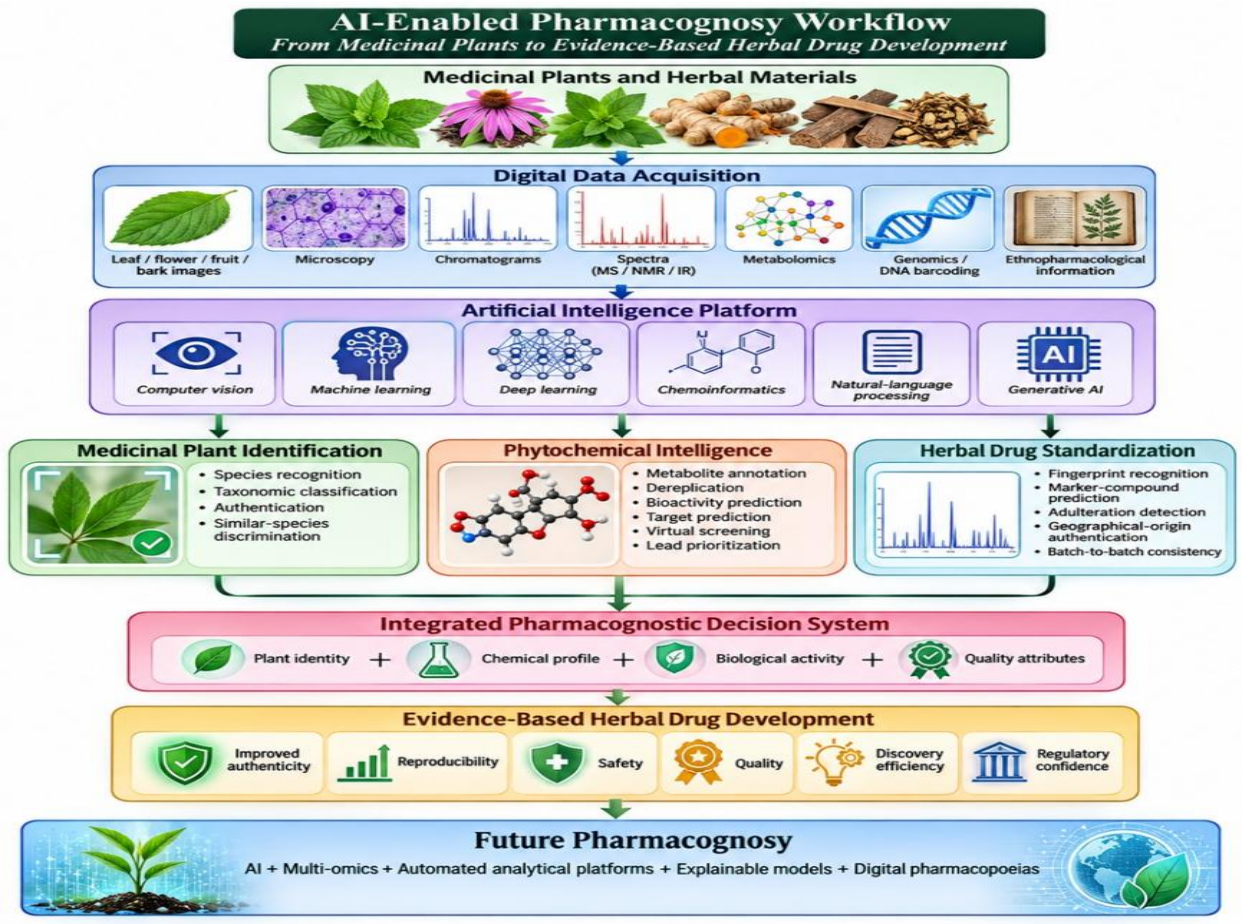

Figure 1. Graphical Abstract

Abstract—**Pharmacognosy is undergoing a major technological transition as artificial intelligence (AI), machine learning (ML), deep learning (DL), computer vision, chemoinformatics and multi-omics become increasingly integrated into the study of medicinal plants and natural products. Conventional pharmacognostic workflows remain indispensable, but they are often dependent on expert taxonomic expertise, labour-intensive microscopic and macroscopic evaluation, extensive phytochemical screening, and analytical procedures that may require substantial time and resources. AI provides an opportunity to transform these workflows by converting complex visual, chemical, genomic and analytical datasets into predictive and decision-support systems. In medicinal plant identification, computer-vision algorithms can recognize species from leaf, flower, fruit, bark and microscopic characteristics, potentially reducing taxonomic errors and improving authentication. Recent deep-learning studies have demonstrated highly accurate classification of medicinal plants from leaf images, although performance can decline when models encounter field conditions, variable illumination, developmental stages or closely related species. AI is also reshaping phytochemical discovery through prediction of bioactivity, molecular target interactions, metabolite annotation, dereplication, virtual screening, compound optimization and integration of metabolomics with genomics and other omics datasets. Recent reviews emphasize that AI-assisted natural-product research can accelerate the prioritization of chemically and biologically promising compounds while reducing dependence on exhaustive experimental screening. In herbal drug standardization, ML coupled with chromatographic, spectroscopic and metabolomic fingerprints can support authentication, geographical-origin determination, detection of adulteration and prediction of quality attributes. A recent metabolomics–ML study involving turmeric and ashwagandha reported 98% specificity and accuracy for authentication-related classifications, illustrating the practical potential of this approach. This review critically examines the emerging role of AI across the pharmacognosy research continuum, from medicinal plant identification and phytochemical discovery to herbal drug quality assessment and standardization. Particular emphasis is placed on computational methodologies, data requirements, explainability, validation, regulatory considerations and future integration with multi-omics and foundation-model technologies.**

Index Terms—**Artificial intelligence; Pharmacognosy; Machine learning; Deep learning; Medicinal plant identification; Phytochemical discovery; Herbal drug standardization; Metabolomics**

I. INTRODUCTION

Medicinal plants represent one of the oldest sources of therapeutic knowledge and continue to provide chemically diverse molecules for modern pharmaceutical research. Natural products have historically contributed directly or indirectly to numerous therapeutically important agents, while contemporary natural-product research continues to explore chemical space that is difficult to reproduce through conventional synthetic approaches. Nevertheless, the discovery and development of medicinal-plant-derived therapeutics remain constrained by several interconnected challenges, including correct botanical identification, chemical complexity, seasonal and geographical variation, adulteration, incomplete metabolite characterization, inefficient screening and difficulties in translating traditional-use information into reproducible pharmaceutical evidence. AI is increasingly being investigated as a means of addressing these limitations by integrating heterogeneous datasets and identifying patterns that may be difficult to recognize through conventional analytical workflows. [1]

Pharmacognosy traditionally depends on a combination of organoleptic, macroscopic, microscopic, physicochemical, chromatographic and phytochemical examinations. These approaches remain foundational because botanical identity and herbal quality cannot be reduced entirely to computational predictions. However, many of these procedures require experienced personnel and can become difficult to scale when thousands of plant samples, images or analytical fingerprints must be evaluated. The development of computer vision and deep-learning methods has therefore introduced an alternative strategy in which morphological information can be transformed into numerical representations and classified using trained neural networks. Recent medicinal-plant identification studies demonstrate the feasibility of this approach, including a study using 5,878 images representing 30 medicinal species in which DenseNet201 achieved 99.64% accuracy under the evaluated dataset conditions. [2]

The significance of AI in pharmacognosy extends beyond visual identification. Natural-product extracts contain complex mixtures in which biologically important constituents may occur at low

concentrations alongside structurally related metabolites. Conventional bioassay-guided fractionation and analytical dereplication remain essential but may require considerable experimental effort. AI-assisted analysis can support prioritization by combining chemical structures, spectral information, biological activity data, molecular targets and biosynthetic information. Machine-learning models can consequently function as hypothesis-generation systems, directing experimental researchers toward compounds, fractions or plant sources with higher predicted value. [3]

Recent developments have further strengthened this concept through integration of AI with metabolomics, genomics, transcriptomics and proteomics. Such multi-omics approaches can connect plant genotype, biosynthetic pathways, metabolite production and biological responses within a common analytical framework. A 2026 review highlighted the emerging synergy between AI and multi-omics in natural-product research, including applications in molecular representation, binding prediction, drug repurposing, retrosynthesis, compound optimization and quality control. [4]

Herbal drug standardization represents another area in which AI may provide substantial value. Herbal medicines are intrinsically variable biological materials, and their quality can be affected by species, plant part, geographical origin, cultivation conditions, harvesting stage, storage, processing and adulteration. International quality-control frameworks emphasize identity, purity, quality and safety as essential components of herbal-material assessment. AI does not replace these requirements; instead, it can provide computational tools for interpreting the increasingly complex datasets generated by chromatographic, spectroscopic and molecular analyses. The emerging field should therefore not be conceptualized as "AI replacing pharmacognosy." Rather, AI should be regarded as an enabling technology that strengthens the observational, analytical and predictive capabilities of pharmacognosy. This distinction is particularly important because computational performance does not automatically establish botanical identity, chemical authenticity or therapeutic efficacy. A model may achieve high internal accuracy while failing when exposed to samples originating from another geographical region, another season, another laboratory or another instrument. The central scientific challenge is consequently not simply to construct highly accurate models, but to establish generalizable, interpretable, experimentally validated and regulatory-compatible AI systems.

This review examines AI across three interconnected domains of contemporary pharmacognosy: medicinal plant identification, phytochemical discovery and herbal drug standardization. It further evaluates the computational approaches underlying these applications, emerging integration with multi-omics and foundation models, practical limitations, regulatory challenges and future opportunities for human AI collaboration.

II. BACKGROUND AND CONCEPTUAL EVOLUTION OF AI IN PHARMACOGNOSY

2.1 From Classical Pharmacognosy to Computational Pharmacognosy

Classical pharmacognosy developed around direct observation and experimental characterization of natural materials. Botanical taxonomy, morphology, microscopy, physicochemical testing and chemical analysis collectively established the identity and quality of crude drugs. With the introduction of chromatographic and spectroscopic technologies, pharmacognostic investigations progressively shifted from primarily descriptive characterization toward molecular fingerprinting and quantitative chemical profiling. The emergence of chemometrics provided an important conceptual bridge between analytical pharmacognosy and AI. Multivariate statistical methods demonstrated that complex chromatographic or spectroscopic datasets could contain sufficient information to distinguish samples according to species, geographical origin, processing conditions or chemical composition. Machine learning represents a broader evolution of this principle because algorithms can learn complex nonlinear relationships between measured features and predefined outcomes. [5]

The transition becomes particularly important for herbal medicines because their identity is multidimensional. A single morphological feature may not be sufficient for authentication, while a single marker compound may not adequately represent the chemical complexity of an extract. An AI system can potentially combine morphological, microscopic, chemical and molecular information to generate a more comprehensive identity profile.

2.2 Major AI Technologies Relevant to Pharmacognosy

AI is a broad technological category encompassing multiple computational approaches. Within pharmacognosy, several methods are particularly relevant. Machine learning includes supervised algorithms such as support vector machines, random forests, gradient-boosting methods and artificial neural networks. These systems learn relationships between input features and known classes or outcomes. Deep learning uses multilayer neural architectures capable of learning hierarchical representations directly from high-dimensional data. Convolutional neural networks are particularly important for plant-image analysis because they can identify spatial patterns associated with leaf shape, venation, texture and other morphological characteristics. [6] Computer vision converts photographs or microscopic images into computational information. Its potential applications include plant species identification, disease recognition, morphological assessment and authentication of crude botanical materials.

Chemoinformatics enables computational representation and analysis of chemical structures, molecular fingerprints, descriptors and biological activity relationships. Within natural-product research, these methods can support compound classification, similarity searching, virtual screening and lead prioritization. [7] Natural-language processing may facilitate extraction of pharmacognostic information from ethnobotanical literature, patents, traditional-medicine records and scientific publications. Properly structured text-mining systems could help connect plant species, traditional uses, phytochemicals, molecular targets and pharmacological evidence.

Generative AI represents a newer direction in which models generate molecular structures or optimize chemical entities according to predefined objectives. Recent work has expanded this concept toward pseudo-natural products and natural-product-inspired chemical space. [8]

2.3 Why Pharmacognosy Is Particularly Suitable for AI Integration

Pharmacognosy generates unusually heterogeneous datasets. A single medicinal plant can be represented by its botanical identity, geographical origin, morphology, microscopic characters, DNA sequence, metabolomic fingerprint, isolated constituents, pharmacological activities, traditional uses and toxicity profile. These datasets differ substantially in format and scale.

AI is particularly valuable when multiple data modalities must be integrated. Instead of treating morphology, chemistry and pharmacology as isolated datasets, multimodal computational frameworks can potentially establish relationships between them. This could allow researchers to ask higher-order questions such as whether a particular plant phenotype predicts a characteristic metabolomic profile, whether metabolomic signatures can predict biological activity, or whether specific biosynthetic genes explain geographical variation in bioactive metabolites.

The growing availability of digital analytical data has therefore created the possibility of a digital pharmacognosy ecosystem, in which plant identity, chemical composition and biological activity are linked computationally. However, realizing this concept requires careful attention to metadata, dataset standardization, reproducibility and experimental validation.

III. ARTIFICIAL INTELLIGENCE FOR MEDICINAL PLANT IDENTIFICATION

3.1 The Need for Automated Identification

Correct botanical identification is the foundation of pharmacognostic research. Misidentification can compromise phytochemical investigations, pharmacological experiments, toxicological evaluations and ultimately herbal-product safety. The problem is especially relevant when morphologically similar species share vernacular names or when medicinal materials are supplied in fragmented, powdered or processed forms. Traditional identification relies heavily on taxonomic expertise. Herbarium comparison, dichotomous keys and microscopic examination remain authoritative approaches, but they may be difficult to implement rapidly in field environments or high-throughput screening programs. AI-based image recognition has emerged as a complementary approach. A deep-learning system can be trained using labelled images of medicinal plants. During training, the algorithm learns visual features associated with different species. When a new image is presented, the trained model

calculates the probability that the image belongs to each learned category.

3.2 Computer Vision and Deep Learning

Convolutional neural networks have become particularly influential in automated plant identification. Their architecture enables progressive extraction of low-level and high-level visual features, beginning with edges and textures and progressing toward more complex morphological patterns. Transfer learning can further reduce the amount of training data required by adapting neural networks previously trained on large image datasets.

Research on medicinal plants has demonstrated the feasibility of deep-learning-based classification. A comparative investigation of seven convolutional neural-network architectures using 5,878 images from 30 medicinal species found substantial classification performance, with DenseNet201 achieving 99.64% accuracy in one evaluated dataset configuration. [9] Another study using an ensemble of deep-learning models similarly demonstrated the potential of combining transfer-learning architectures for automatic medicinal-leaf identification. [10]

However, numerical accuracy should not be interpreted without considering dataset composition. Image-based models may learn background characteristics, lighting conditions or photographic artifacts rather than genuinely taxonomic features. Performance obtained from controlled images may therefore overestimate performance in natural environments.

3.3 Beyond Leaves: Multimodal Botanical Identification

Leaves are convenient because they are accessible and visually informative, but reliable pharmacognostic identification cannot always depend on a single plant organ. Closely related species may exhibit highly similar leaves while differing in flowers, fruits, bark, seeds, roots or microscopic characters.

3.4 AI-Assisted Identification of Crude and Processed Drugs

A major challenge arises when medicinal plants are supplied as dried pieces, powders or extracts. Morphological information becomes progressively less available as processing increases. Here, AI can shift from conventional image recognition toward microscopic image analysis and analytical fingerprint recognition. Microscopy combined with image-analysis algorithms may identify characteristic cell structures, vascular patterns, trichomes, starch grains, crystals or other diagnostic features. Similarly, chromatographic or spectroscopic fingerprints can be treated as high-dimensional patterns that ML algorithms classify according to authenticated reference materials. This approach has particular relevance to quality-control laboratories because computational classification can support rapid screening before confirmatory testing. It should nevertheless remain a decision-support tool rather than an autonomous replacement for pharmacognostic authentication.

3.5 Challenges in AI-Based Plant Identification

Several factors limit generalizability:

1. Limited training datasets for medicinal species.
2. Class imbalance, where some species have thousands of images while others have very few.
3. Inter-species similarity, particularly among related taxa.
4. Intra-species variability caused by age, season and environmental conditions.
5. Image-quality variation involving illumination, camera type and resolution.
6. Background bias in field photographs.
7. Geographical bias resulting from datasets collected from restricted locations.
8. Taxonomic uncertainty in source labels.
9. Lack of external validation using independent datasets.
10. Limited explainability regarding why the model selected a particular species.

The most scientifically robust strategy is therefore likely to combine AI classification with expert taxonomic confirmation and independent laboratory evidence.

Table 1. Major Artificial Intelligence Technologies and Their Applications in Pharmacognosy [11]

AI technology	Principal data type	Major pharmacognostic application	Potential advantage	Important limitation
Machine learning	Tabular, chemical and analytical data	Classification, authentication and prediction	Flexible and relatively interpretable algorithms	Dependent on feature quality
Deep learning	Images, spectra, large datasets	Medicinal plant identification and complex pattern recognition	Automated feature extraction	Requires substantial representative datasets
Computer vision	Plant and microscopic images	Species identification and crude-drug authentication	Rapid non-destructive screening	Sensitive to image and environmental variability
Chemoinformatics	Molecular structures and descriptors	Compound prioritization and virtual screening	Enables large-scale chemical analysis	Prediction may not translate directly to biological activity
Natural-language processing	Literature and ethnobotanical records	Knowledge extraction and evidence mapping	Converts unstructured literature into searchable information	Ambiguity and inconsistent terminology
Generative AI	Molecular representations	Natural-product-inspired compound generation and optimization	Expands searchable chemical space	Generated molecules require experimental validation
Multi-omics AI	Genomics, transcriptomics, proteomics and metabolomics	Biosynthetic pathway analysis and integrated discovery	Links genotype, chemistry and phenotype	High dimensionality and data-standardization challenges

IV. AI FOR PHYTOCHEMICAL DISCOVERY

Phytochemical discovery has traditionally involved extraction, fractionation, chromatographic separation, spectroscopic characterization and biological evaluation. Although these methods remain fundamental, the chemical complexity of plant extracts can make the discovery process inefficient. AI offers opportunities to prioritize compounds and samples before extensive experimental investigation.

4.1 AI-Assisted Dereplication

Dereplication is one of the most valuable applications of computational methods in natural-product research. Its objective is to recognize known compounds rapidly so that experimental resources can be directed toward potentially novel metabolites.

AI can assist by comparing mass-spectral, fragmentation and structural information against large chemical databases. Machine-learning approaches can learn relationships among molecular features and facilitate annotation of unknown or partially characterized metabolites. This is especially relevant to natural-product extracts because thousands of molecular signals may be detected in a single metabolomic experiment.

The integration of AI with computational natural-product databases is increasingly transforming the workflow from sequential identification toward prioritization-driven discovery. [12]

4.2 Prediction of Biological Activity

A central objective of phytochemical research is identifying compounds with desirable biological activity. AI models can learn associations between chemical structures and experimental activity data. Depending on the available dataset, models may predict antimicrobial, antioxidant, anti-inflammatory, anticancer, antiviral or other pharmacological properties.

Such predictions should be interpreted as hypotheses rather than experimental evidence. A compound predicted to possess strong activity may fail because of poor bioavailability, instability, inadequate cellular exposure, protein-binding limitations or mechanisms not represented in the training dataset.

Nevertheless, predictive modelling can reduce the number of compounds requiring initial experimental

screening and can help researchers prioritize chemically diverse candidates.

4.3 Molecular Target Prediction and Network Pharmacology

Natural products frequently act through multiple targets rather than a single molecular pathway. This characteristic complicates traditional target-identification approaches but simultaneously creates opportunities for network-based computational analysis. AI can integrate chemical structures with known target interactions, gene-expression data, protein networks and disease-associated pathways. The resulting systems can generate hypotheses regarding potential molecular targets and biological mechanisms. The combination of AI and network pharmacology is particularly relevant to traditional herbal medicines because therapeutic effects may arise from multiple constituents acting across interconnected pathways. Recent natural-product research has highlighted AI-assisted mechanistic analysis as an emerging component of this field. [13]

4.4 AI and Biosynthetic Pathway Prediction

Plants possess sophisticated biosynthetic machinery capable of generating structurally diverse secondary metabolites. Genomic and transcriptomic information can provide clues regarding enzymes responsible for specialized metabolite biosynthesis.

AI and machine-learning approaches can integrate sequence information with metabolomic observations to identify candidate biosynthetic genes or enzymes. Such systems may eventually enable prediction of which plant species, tissues or environmental conditions are most likely to produce a particular class of bioactive metabolites.

This represents an important transition from discovering compounds after extraction toward predicting where chemically valuable compounds are likely to occur before extensive isolation work begins.

4.5 Generative AI and Natural-Product-Inspired Molecules

Generative models extend AI beyond prediction toward molecular creation. Deep generative systems can learn patterns within chemical datasets and generate new structures according to desired molecular properties. In natural-product research, this concept is particularly attractive because natural products possess complex stereochemical and scaffold diversity that can inspire new chemical entities.

Recent work has explored machine-learning-guided generation of pseudo-natural products, although challenges remain regarding chemical novelty, synthetic feasibility, drug-like properties and biological validation. [14]

The emerging objective is therefore not merely to reproduce naturally occurring molecules but to use natural-product chemical space as a design template for discovering compounds with optimized pharmacological characteristics.

Table 2. AI Applications Across the Phytochemical Discovery Pipeline [15]

Discovery stage	Conventional approach	AI-enabled approach	Expected contribution
Plant/source selection	Ethnobotanical and literature-based selection	Predictive prioritization using biological and chemical datasets	Improved selection of promising sources
Extract profiling	Chromatography and spectroscopy	ML-assisted fingerprint classification	Rapid pattern recognition
Dereplication	Manual database comparison	Automated spectral and structural matching	Reduced rediscovery of known compounds
Metabolite annotation	Expert interpretation	AI-assisted spectral prediction and annotation	Faster identification
Bioactivity screening	Experimental screening of large libraries	Activity prediction and prioritization	Reduced experimental burden
Target identification	Literature and experimental methods	Structure–target prediction and network analysis	Mechanistic hypothesis generation
Lead optimization	Conventional medicinal chemistry	Generative AI and predictive modelling	Optimization of molecular properties
Biosynthetic investigation	Experimental pathway characterization	Multi-omics prediction	Identification of candidate pathways and enzymes

V. AI-DRIVEN HERBAL DRUG STANDARDIZATION

Herbal drug standardization is arguably one of the areas in which AI can provide the greatest practical benefit because herbal materials are chemically and biologically heterogeneous. International guidance emphasizes quality control of medicinal plant materials and appropriate selection of marker substances for herbal medicines.

5.1 From Single Markers to Multidimensional Fingerprints

Conventional standardization frequently emphasizes marker compounds. Although marker-based testing remains useful, a single constituent may not adequately capture the chemical identity of a complex herbal preparation.

Chromatographic fingerprints, mass-spectral profiles and metabolomic datasets provide a broader representation of herbal-material composition. ML algorithms can analyze these multidimensional patterns and distinguish authentic materials from adulterated or chemically divergent samples.

5.2 Detection of Adulteration

Adulteration is a major quality and safety concern. Substitution may occur deliberately for economic reasons or unintentionally because morphologically similar species are confused during collection and processing.

AI-assisted chromatographic and spectroscopic analysis can identify subtle differences that may be difficult to evaluate through visual inspection alone. ML classifiers can be trained using authenticated and adulterated samples and subsequently used to flag unknown materials.

A recent study integrating LC-Orbitrap high-resolution mass spectrometry with untargeted metabolomics and machine learning investigated turmeric and ashwagandha authentication across geographical origin, variety, tissue specificity and adulteration. The reported models achieved approximately 98% specificity and accuracy for the evaluated classification tasks, demonstrating how analytical chemistry and AI can be integrated for herbal authentication. [16]

5.3 Geographical-Origin Authentication

The chemical composition of medicinal plants can vary according to soil, climate, altitude, rainfall, temperature and cultivation practices. Consequently, geographical origin can influence both quality and therapeutic consistency.

AI can identify subtle metabolomic signatures associated with geographic origin. Such models could become valuable for traceability and quality assurance, particularly for medicinal plants with high commercial value or protected geographical identities. However, geographical models must be externally validated because environmental conditions may change over time and overlapping chemical profiles may occur across regions.

5.4 Batch-to-Batch Consistency

Commercial herbal products may be manufactured from different harvests, suppliers or geographic sources. AI can compare analytical fingerprints across production batches and identify deviations from predefined reference profiles.

A validated computational system could potentially classify batches into acceptable and unacceptable categories or provide a continuous similarity score. This could complement established quality-control testing and facilitate earlier detection of process deviations.

5.5 Contaminant and Safety Assessment

Herbal materials may contain pesticides, heavy metals, microbial contaminants, mycotoxins or undeclared pharmaceutical ingredients. AI could assist in identifying abnormal analytical signatures associated with contamination.

Nevertheless, safety-critical decisions require validated analytical methods. AI should not be used as an independent substitute for established contaminant testing, particularly when regulatory thresholds are involved.

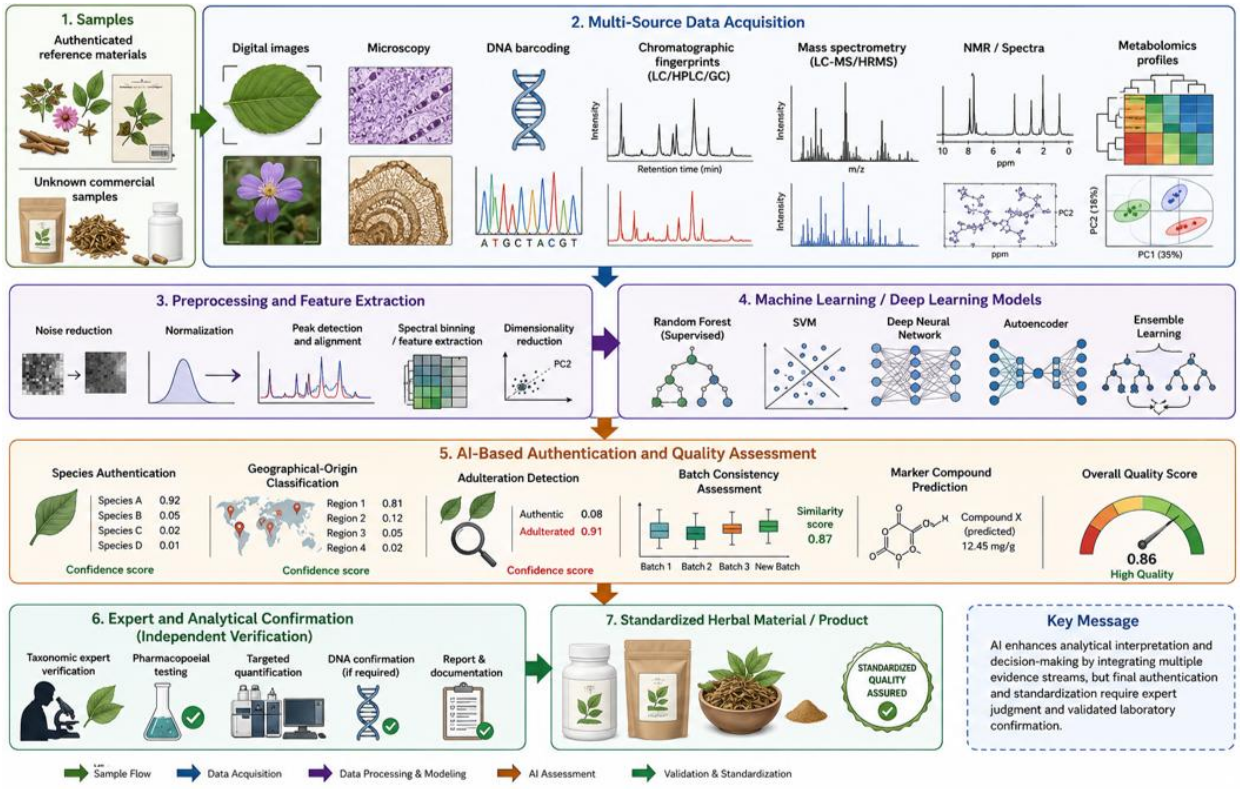

Figure 2. AI-Enabled Medicinal Plant Identification [17]

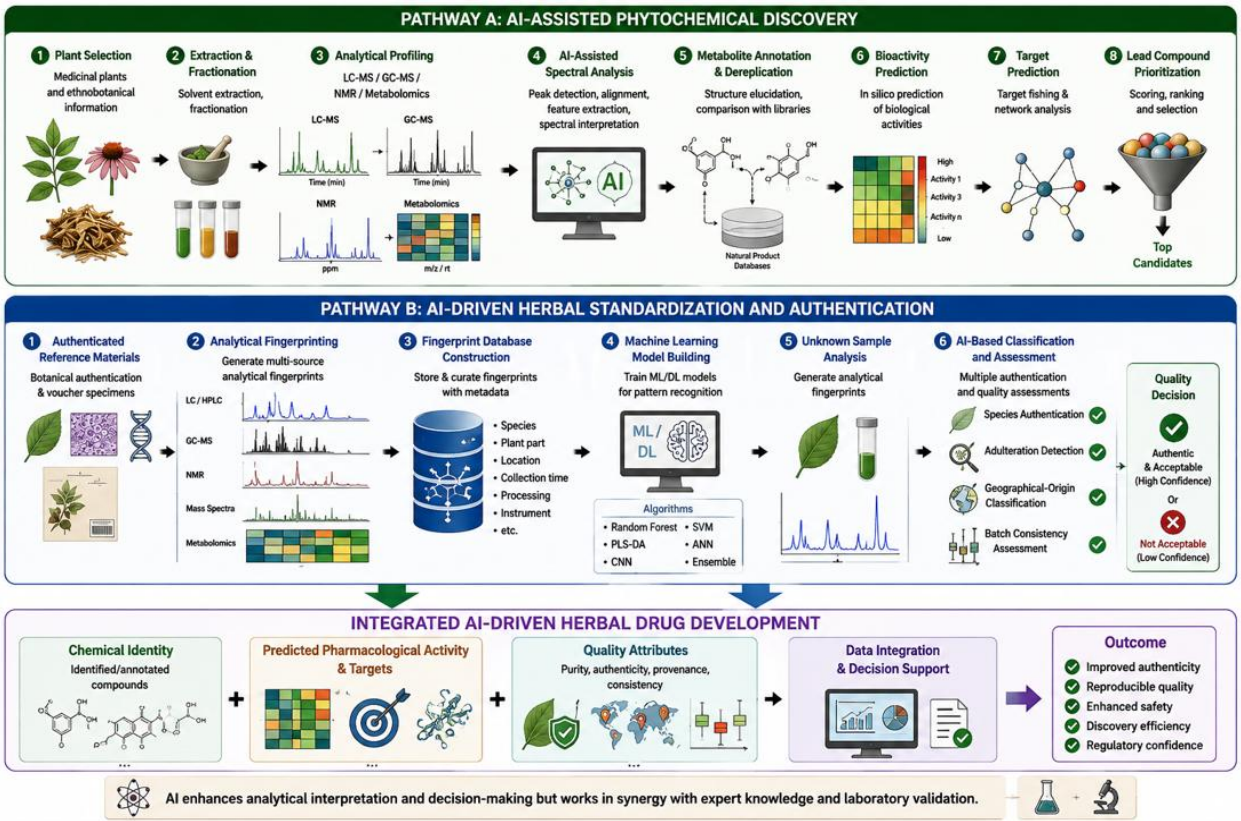

Figure 3. AI-Assisted Phytochemical Discovery and Herbal Standardization [18]

VI. MOLECULAR AND COMPUTATIONAL MECHANISMS UNDERLYING AI-ENABLED PHARMACOGNOSY

The application of artificial intelligence in pharmacognosy is fundamentally dependent on the conversion of complex biological observations into standardized digital representations. Medicinal plants can generate multiple forms of information, including macroscopic images, microscopic characters, DNA sequences, chromatographic fingerprints, mass spectra, metabolomic profiles and pharmacological measurements. AI provides computational frameworks capable of learning relationships within and between these data types, thereby supporting classification, prediction and prioritization. The increasing convergence of AI with natural-product research is particularly significant because natural-product discovery traditionally involves chemically complex mixtures, extensive experimental screening and substantial problems of rediscovery. [19]

6.1 Data Acquisition and Digital Representation

The first stage of an AI-enabled pharmacognostic workflow is the generation of reliable and appropriately annotated data. For medicinal-plant identification, this may involve standardized photographs of leaves, flowers, fruits, stems, roots or whole plants. Microscopic studies can generate digital images of diagnostic anatomical structures. Chemical investigations can produce chromatograms, mass spectra, NMR spectra and metabolomic datasets.

The quality of these inputs directly affects model reliability. Differences in sample preparation, instrument configuration, image resolution, extraction procedure or geographical origin can introduce variation that an algorithm may incorrectly interpret as a biologically meaningful characteristic. Consequently, AI datasets should include detailed metadata describing botanical identity, plant part, collection location, collection period, processing conditions and analytical methodology. Recent reviews of machine-learning applications in medicinal plants emphasize that data quality, standardization and validation remain central challenges for successful implementation. [20]

6.2 Feature Extraction and Representation Learning

Conventional machine-learning approaches often require researchers to identify informative features before model construction. In botanical identification, these features may include leaf dimensions, perimeter, shape, venation, texture or color distribution. Deep-learning systems differ because they can learn hierarchical representations directly from images.

Convolutional neural networks can progressively transform raw pixels into increasingly complex feature representations. Early layers may identify edges and textures, whereas deeper layers can recognize combinations of morphological characteristics. This capability has made deep learning particularly attractive for medicinal-plant recognition.

A study evaluating seven deep-learning architectures using 5,878 images representing 30 medicinal species reported 99.64% accuracy for DenseNet201 under its public-image dataset and 97% accuracy when field images with more complex backgrounds were included. The difference between these conditions illustrates why model performance must be evaluated using datasets that reflect the intended real-world environment. [21]

6.3 Classification, Regression and Clustering

AI applications in pharmacognosy can be broadly categorized according to the type of computational problem being addressed.

Classification assigns samples to predefined categories. Examples include species identification, authentic-versus-adulterated classification and geographical-origin determination.

Regression predicts continuous variables such as marker-compound concentration, quality scores or biological activity.

Clustering identifies naturally occurring groups without requiring predefined labels. This approach is particularly useful for metabolomics because chemically similar samples may form clusters corresponding to chemotypes, geographical populations or processing conditions.

The appropriate algorithm should therefore be selected according to the scientific objective, dataset characteristics and validation requirements rather than according to algorithmic complexity alone. [22]

6.4 Training, Validation and External Testing

A scientifically reliable AI model requires separation of data into training, validation and independent testing datasets. The training dataset is used for model development, validation data support optimization and

model selection, while an independent test dataset estimates performance on previously unseen samples.
A particularly important concern in natural-product research is data leakage. Samples originating from the same plant population, extraction batch or analytical run may be highly similar. If related observations are distributed across both training and test datasets, the reported performance may be artificially inflated.
Biologically meaningful splitting strategies should therefore be used. For example, samples from a particular plant population or production batch may need to be kept entirely outside the training dataset when evaluating generalization to new populations or batches.
External validation is especially important before AI models are used for herbal authentication or quality-control decisions. Recent natural-product AI reviews specifically identify dataset quality and appropriate algorithm validation as major requirements for successful translation. [23]

VII. INTEGRATION OF AI WITH METABOLOMICS AND MULTI-OMICS

7.1 Metabolomics as a Foundation for AI-Driven Pharmacognosy

Metabolomics provides a broad representation of the small molecules present within biological systems. This is particularly valuable for medicinal plants because their chemical phenotype can contain hundreds or thousands of detectable metabolites.
Untargeted metabolomics can therefore generate a chemical fingerprint that extends beyond conventional marker-compound analysis. AI and machine-learning approaches can identify patterns within these high-dimensional datasets and associate them with species, tissue type, geographical origin, environmental conditions or pharmacological properties.
The integration of metabolomics with AI is increasingly recognized as a powerful strategy for natural-product discovery because it can help prioritize chemically interesting samples and facilitate dereplication. [24]

7.2 Genomics–Metabolomics Integration

Genomic information provides insight into the biosynthetic potential of a medicinal plant, whereas metabolomics describes the chemical products generated by that biosynthetic machinery.
AI can potentially connect these two levels by identifying statistical associations between genes and metabolites. Such relationships may help researchers identify candidate biosynthetic genes responsible for producing pharmacologically important compounds.
This approach could substantially change natural-product discovery. Instead of beginning with random extraction and isolation, researchers could use genomic and metabolomic information to identify plant tissues or populations predicted to contain desirable chemical classes.
Recent research on plant natural-product biosynthesis emphasizes the increasing role of big-data analytics, machine learning and computational methods in identifying regulatory networks and components of specialized metabolic pathways. [25]

7.3 Transcriptomics and Phytochemical Biosynthesis

Transcriptomics provides information about gene-expression patterns under specific developmental or environmental conditions. Medicinal plants can substantially modify their secondary metabolism in response to drought, temperature, pathogens, nutrient availability and other environmental stresses.
AI-based integration of transcriptomic and metabolomic datasets may reveal associations between gene-expression patterns and accumulation of particular phytochemicals. This could assist in identifying candidate regulatory genes and biosynthetic pathways.
Such approaches may also contribute to precision cultivation by identifying environmental conditions associated with desirable phytochemical profiles.

7.4 Proteomics and Biosynthetic Enzymes

Proteomics adds another layer of information by characterizing proteins and enzymes involved in metabolic processes. When proteomic data are combined with transcriptomics and metabolomics, researchers can move closer to reconstructing functional biosynthetic pathways.
AI is particularly useful because these datasets are high-dimensional and contain complex relationships that may be difficult to evaluate through conventional univariate methods. Recent multi-omics literature describes AI and machine learning as important tools for integrating gene, metabolite and protein information in natural-product research. [26]

7.5 AI as an Integrative Layer

A conceptual multi-omics workflow can be represented as:

Genome → Transcriptome → Proteome → Metabolome → Phenotype → Pharmacological activity

The purpose of AI is not simply to analyze each layer separately, but to identify relationships across these biological levels.

The emerging combination of AI and multi-omics has been proposed for target identification, natural-product characterization, molecular binding prediction, compound optimization, quality control and drug-development applications. A recent 2026 review specifically highlighted AI-assisted integration of genomics, transcriptomics, proteomics and metabolomics in natural-product research. [27]

Table 3. Multi-Omics Data Integration in AI-Enabled Pharmacognosy [28]

Omics layer	Primary information	Pharmacognostic relevance	Potential AI application
Genomics	Genes, variants and biosynthetic potential	Species characterization and biosynthetic capacity	Gene-function and pathway prediction
Transcriptomics	Gene-expression patterns	Regulation of secondary metabolism	Gene–metabolite association
Proteomics	Proteins and enzymes	Functional biosynthetic machinery	Enzyme and pathway prediction
Metabolomics	Small-molecule profiles	Chemical phenotype and authentication	Classification and biomarker discovery
Phenotypic data	Morphological and physiological characteristics	Botanical identity and environmental response	Image classification and phenotype prediction
Pharmacological data	Biological activity	Therapeutic relevance	Activity and target prediction

VIII. AI, NETWORK PHARMACOLOGY AND MECHANISTIC INTERPRETATION

Medicinal plants frequently contain numerous constituents capable of interacting with multiple molecular targets. This polypharmacological characteristic creates challenges for conventional single-target approaches but also provides an opportunity for network-based computational analysis. Network pharmacology represents relationships among compounds, targets, proteins, pathways and disease phenotypes. AI can extend this framework by learning patterns from chemical–target and disease-associated datasets.

A conceptual workflow can therefore be represented as:

Phytochemical structure → predicted target → protein interaction network → signaling pathway → disease phenotype

Such computational frameworks can generate mechanistic hypotheses for complex herbal preparations and prioritize molecular interactions for experimental investigation. AI-assisted target prediction is particularly promising when combined with natural-product chemical information. However, prediction quality depends on whether the training data adequately represent natural-product chemical space. Natural products frequently contain complex scaffolds and stereochemical arrangements that may be poorly represented in conventional synthetic-drug datasets. [29] The use of AI in network pharmacology should therefore be regarded as a hypothesis-generation strategy, followed by molecular, cellular and pharmacological validation

IX. ARTIFICIAL INTELLIGENCE FOR HERBAL AUTHENTICATION AND ADULTERATION DETECTION

9.1 Botanical Authentication

Botanical authentication represents the first major quality requirement for medicinal plant materials. AI-based image classification can compare unknown plant images with authenticated reference datasets and generate probable species classifications.

Deep-learning systems have already demonstrated strong performance in medicinal-leaf recognition. However, controlled laboratory images and real-world field images can produce substantially different model performance, emphasizing the importance of environmental diversity during model development. [30]

The most reliable future systems are likely to combine AI-based visual classification with independent pharmacognostic evidence. For difficult cases, the workflow could become:
Image recognition → microscopic examination → DNA authentication → chemical fingerprint → integrated identity decision.
This multimodal approach would reduce the risk associated with relying on one source of evidence.

9.2 Chemical Authentication
Chromatographic and spectroscopic fingerprints contain multidimensional information regarding the chemical composition of herbal materials. Machine-learning algorithms can analyze these patterns and distinguish authenticated materials from chemically divergent samples. This approach represents an important shift from single-marker testing toward holistic chemical characterization. Chemometric and machine-learning methods have been increasingly recognized as useful tools for botanical authentication because they can interpret complex small-molecule profiles and integrate information from multiple analytical techniques. [31]

9.3 Adulteration Detection
Adulteration may involve substitution with another plant species, use of an incorrect plant part, dilution with cheaper material or addition of undeclared substances. Because adulterated materials may retain some expected marker compounds, single-marker analysis may not always detect subtle substitution.
AI can instead evaluate the complete analytical pattern. Machine-learning models trained with authentic and adulterated samples can identify distributed chemical differences across multiple features.
A recent LC-Orbitrap-HRAMS metabolomics study of turmeric and ashwagandha demonstrated the practical potential of this approach. Machine-learning models were able to classify geographical origin, variety and tissue specificity and reported 98% specificity and accuracy for the investigated authentication tasks, including adulterated market samples. [32]

9.4 Geographical-Origin Authentication
Geographical origin can influence medicinal-plant chemistry because climate, altitude, soil, rainfall and cultivation conditions affect plant metabolism.
AI can recognize geographical patterns in metabolomic or spectroscopic data and classify samples according to origin. This can support traceability and quality assurance for commercially important botanicals.
However, geographical models require broad sampling because models trained using a small number of locations may learn local characteristics that do not generalize to other production regions. Recent reviews of herbal quality control specifically identify geographical-origin tracing as a major application of analytical chemistry combined with machine learning. [33]

X. AI-DRIVEN QUALITY CONTROL AND STANDARDIZATION OF HERBAL MEDICINES

10.1 From Single Markers to Chemical Fingerprints
Herbal medicines are complex mixtures, and their quality cannot always be adequately represented by one marker compound. Fingerprinting approaches provide a broader description of chemical composition.
AI can analyze chromatographic, spectroscopic and metabolomic fingerprints and determine whether an unknown sample falls within the expected chemical distribution of authenticated reference materials.
A 2026 review of herbal-medicine quality control described the integration of infrared spectroscopy, mass spectrometry and NMR with machine learning for geographical-origin tracing, adulteration detection and species identification. The review emphasized the ability of ML to interpret high-dimensional chemical data and support more holistic quality-assessment frameworks than single-marker analysis alone. [34]

10.2 Quantitative Prediction of Marker Compounds
Machine-learning regression models can potentially estimate concentrations of selected phytochemical markers from spectroscopic or chromatographic information.
This could support rapid preliminary screening of large numbers of samples. However, computationally estimated concentrations should not automatically replace validated quantitative analytical methods.
For regulatory quality control, AI-derived predictions should be appropriately calibrated against reference analytical methods and subjected to predefined acceptance criteria.

10.3 Batch-to-Batch Consistency

Consistency between manufacturing batches is a major requirement for herbal-drug quality. Raw-material variability, extraction conditions, processing parameters and storage can all influence the final chemical profile. AI can compare multivariate fingerprints from new batches against authenticated reference populations. A model may classify batches as conforming or non-conforming or generate a similarity score indicating their position relative to the established reference distribution. This approach could provide an early-warning mechanism for manufacturing deviations.

10.4 Predictive Quality Control

A future quality-control system could connect:

Cultivation conditions → Raw-material chemistry → Extraction parameters → Chemical fingerprint → Finished-product quality

Historical manufacturing datasets could be used to identify which variables most strongly influence final quality. This would shift herbal quality control from predominantly retrospective testing toward predictive and process-oriented quality management.

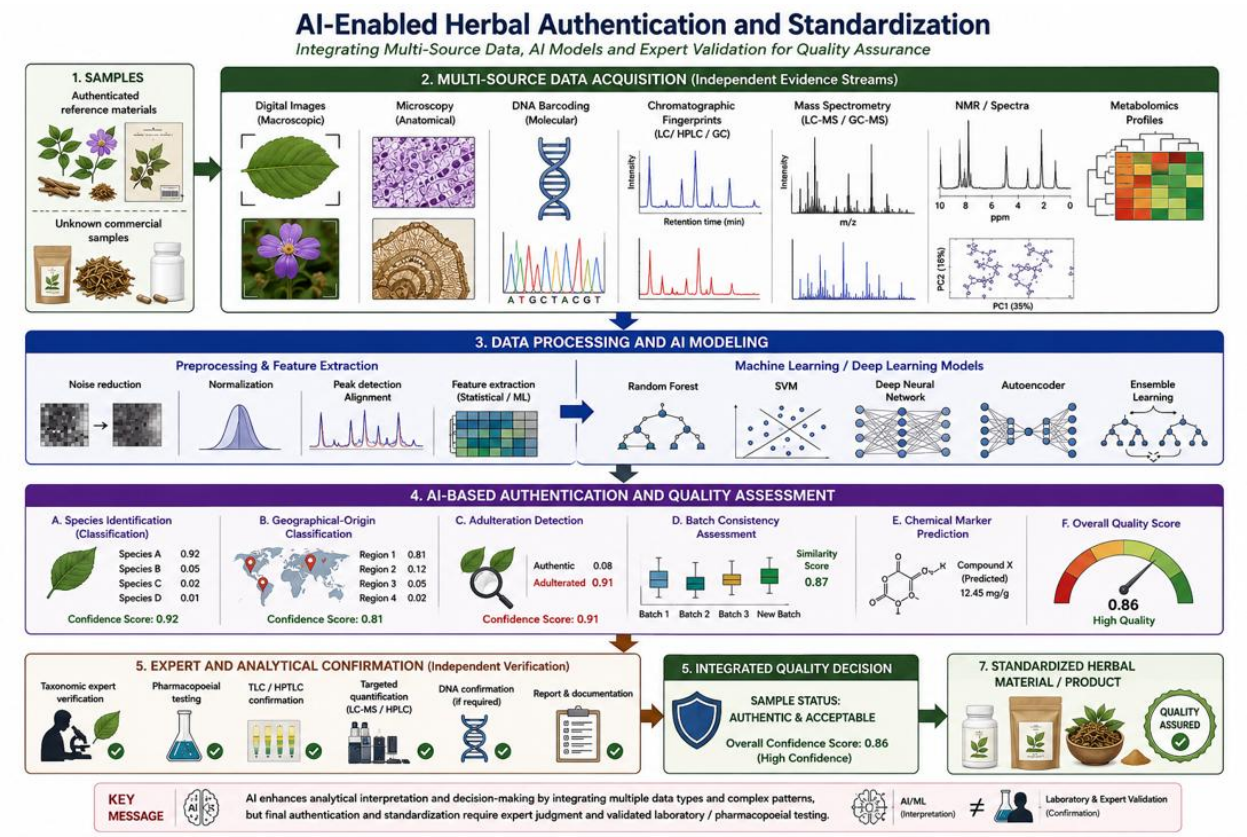

Figure 4. AI-Enabled Herbal Authentication and Standardization [35]

XI. RECENT ADVANCES IN AI-ENABLED PHARMACOGNOSY

11.1 Deep Learning for Medicinal Plant Recognition

The use of deep convolutional neural networks has expanded rapidly in medicinal-plant identification. Comparative studies have demonstrated high classification accuracy under controlled conditions, while real-time identification systems have also been investigated. The major scientific question is now shifting from whether deep learning can identify medicinal plants to whether these models can generalize across different geographic locations, seasons, imaging devices, plant developmental stages and environmental backgrounds. Systematic reviews of deep learning in medicinal-plant recognition have highlighted both the potential of these approaches and

the need for larger, more diverse datasets and stronger validation strategies. [36]

11.2 AI–Metabolomics Integration

The combination of untargeted metabolomics and machine learning is becoming one of the most important developments in computational herbal science. Rather than identifying only predetermined marker compounds, AI can recognize patterns involving numerous metabolites simultaneously. This enables authentication, adulteration detection, geographical classification and potentially prediction of biological activity. The turmeric and ashwagandha study described above provides a recent example of this approach in practice. [37]

11.3 Natural-Product Knowledge Graphs

Natural-product research generates extensive information linking plants, compounds, targets, pathways, traditional uses and pharmacological activities. Knowledge graphs can organize these relationships into machine-readable networks. Natural-language processing can potentially extract relevant information from scientific publications and connect it with existing chemical and biological databases. Such systems could eventually enable researchers to move from a plant name to a structured evidence map containing known compounds, biological targets, traditional uses, experimental evidence and unresolved research questions.

Recent AI reviews have identified knowledge integration and multimodal data analysis as important components of the next generation of natural-product research. [38]

11.4 Foundation Models and Generative AI

Foundation models and generative AI are emerging beyond conventional task-specific machine-learning systems. In natural-product research, generative models may assist with molecular representation, virtual screening, molecular optimization and generation of natural-product-inspired chemical structures. A 2026 review described the expanding role of AI and multi-omics in natural-product drug development, including binding prediction, drug repurposing, retrosynthesis and generation or optimization of pseudo-natural compounds. [39]

However, generated molecular structures remain computational hypotheses. Chemical synthesis, structural confirmation, biological activity, pharmacokinetic properties and toxicity must be experimentally established before therapeutic significance can be assigned.

XII. CLINICAL AND PHARMACEUTICAL TRANSLATION

AI-enabled pharmacognosy ultimately becomes scientifically meaningful when computational predictions contribute to reproducible pharmaceutical development. At the discovery stage, AI can prioritize medicinal plants and phytochemicals. At the preclinical stage, it can support target prediction, activity prioritization and safety assessment. During formulation and manufacturing, AI can contribute to analytical standardization and batch-quality assessment. Nevertheless, computational prediction does not constitute clinical evidence.

A phytochemical predicted to interact with a particular molecular target must still demonstrate activity in appropriate experimental systems. Similarly, an AI-authenticated herbal material must undergo validated quality testing before it can be considered suitable for regulated pharmaceutical use.

The most realistic model is therefore an integrated pipeline:

AI prediction → Experimental verification → Preclinical evaluation → Standardization → Clinical evaluation → Regulatory assessment.

Recent natural-product research emphasizes that AI should accelerate the discovery process while maintaining experimental validation as the foundation for biological and therapeutic conclusions. [40]

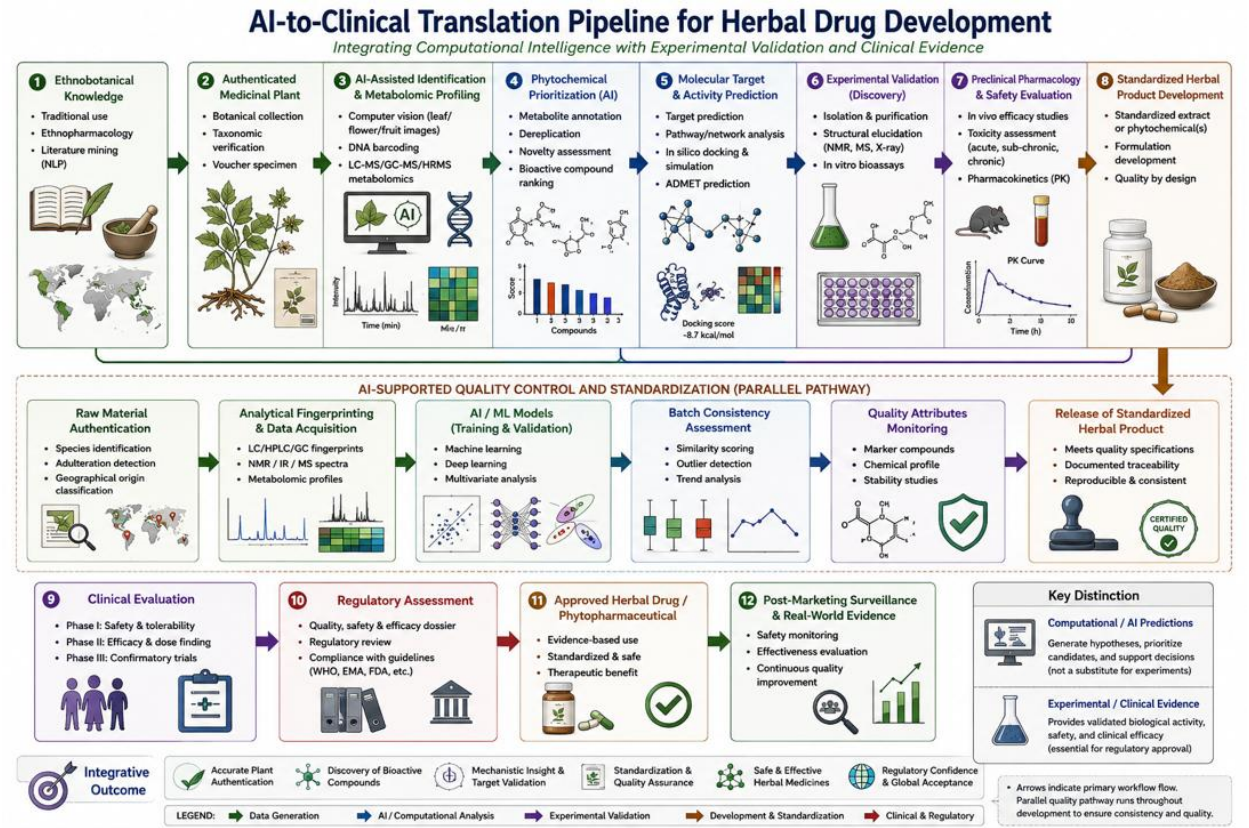

Figure 5. AI-to-Clinical Translation Pipeline for Herbal Drug Development

XIII. CHALLENGES AND LIMITATIONS

Despite considerable technological progress, several barriers must be overcome before AI can become a routine component of pharmacogenetic practice.

Dataset Quality: AI models are strongly influenced by the quality of their training data. Incorrect taxonomic labels, duplicated observations, incomplete metadata and geographically restricted sampling can lead to misleading predictions.

Dataset Bias: A model trained using images from one geographic region may perform poorly in another. Similarly, a chemical model trained using one instrument platform may fail when applied to data generated using another platform.

Limited Natural-Product Data: Many computational drug-discovery datasets contain substantially more information about synthetic pharmaceuticals than natural products. This can limit the ability of AI systems to recognize the unusual structural diversity of phytochemicals.

Interpretability: Deep-learning models may provide highly accurate predictions without clearly explaining the biological or chemical basis for those predictions. Explainability is particularly important when AI is used for botanical authentication or quality-control decisions.

External Validation: High performance on internal datasets does not necessarily indicate real-world reliability. Independent validation across laboratories, geographical regions and sample populations is essential.

Regulatory Acceptance: AI-assisted quality-control systems will require documented model development, validation, version control, performance monitoring and change management before they can be incorporated into regulated workflows.

A recent 2026 review specifically identified data quality, standardization, interpretability and validation as important issues affecting AI adoption in herbal medicine quality control. [41]

Human Expertise: AI cannot independently replace taxonomic expertise, pharmacognostic judgment or laboratory confirmation. Instead, the strongest future systems will likely use AI to identify patterns and prioritize decisions while experts retain responsibility for interpretation and confirmation.

Table 4. Major Challenges and Recommended Strategies for AI Adoption in Pharmacognosy [42]

Challenge	Potential consequence	Recommended strategy
Small datasets	Overfitting and poor generalization	Develop multicenter reference datasets
Incorrect botanical labels	Incorrect learned associations	Expert-verified voucher specimens
Geographic bias	Poor transferability	Include samples from multiple regions
Instrument variability	Analytical batch effects	Standardized acquisition and preprocessing
Black-box models	Limited interpretability	Explainable-AI approaches
Data leakage	Artificially high accuracy	Biologically meaningful data splitting
Limited natural-product data	Poor representation of chemical space	Curated natural-product datasets
Lack of external validation	Uncertain real-world performance	Independent prospective validation
Regulatory uncertainty	Limited routine adoption	Formal AI validation and lifecycle management
Overreliance on prediction	Incorrect biological conclusions	Mandatory experimental confirmation

XIV. FUTURE PERSPECTIVES AND EMERGING RESEARCH DIRECTIONS

The future of pharmacognosy is likely to be shaped not by the replacement of conventional botanical, microscopic and analytical expertise, but by its integration with increasingly sophisticated computational systems. AI can potentially connect information that is currently distributed across separate stages of pharmacognostic research, including botanical identification, chemical characterization, biological evaluation, quality control and clinical translation.

14.1 Multimodal AI and Digital Pharmacognosy

The next generation of AI systems should integrate different evidence types rather than relying on a single image or analytical platform. A medicinal plant may simultaneously generate morphological, microscopic, genomic, metabolomic and pharmacological information. Multimodal models could potentially interpret these datasets together and produce a comprehensive evidence profile.

A future digital pharmacognostic platform could therefore follow the sequence:

Plant specimen → image → microscopy → DNA barcode → metabolome → phytochemical structures → predicted targets → biological evidence → quality profile

Such systems could be particularly valuable when one evidence type is inconclusive. For example, a visual classifier could provide an initial species prediction, while chemical and molecular information could increase or decrease confidence in that classification.

14.2 Development of Large Reference Datasets

The development of reliable AI systems requires large, representative and well-annotated datasets. For medicinal plants, future repositories should ideally include authenticated voucher specimens together with standardized photographs, microscopy, DNA sequences, chromatographic fingerprints, mass spectra and phytochemical information.

The importance of high-quality datasets has been repeatedly emphasized in AI-driven natural-product research because model performance and generalization depend strongly on the quality and diversity of the training data. [43]

Reference datasets should also contain negative examples, closely related species, known adulterants and geographically diverse samples. This would enable models to learn not only what an authentic sample looks like but also what plausible incorrect alternatives look like.

14.3 Explainable AI

Explainability will be essential for adoption in pharmacognostic laboratories. A model should ideally provide more than a probability score. It should indicate the morphological, microscopic or chemical features that contributed to its decision.

For plant-image recognition, visualization techniques could identify regions of an image that strongly

influenced classification. For chemical models, feature-attribution methods could identify spectral regions or molecular fragments associated with a prediction. Explainable systems could improve confidence among pharmacognosists and make model errors easier to identify.

14.4 Federated and Collaborative AI

Individual institutions often possess relatively small datasets. Collaborative approaches could allow universities, botanical institutes, pharmaceutical companies and quality-control laboratories to contribute to model development without necessarily sharing all raw data.

Federated-learning approaches are particularly attractive where datasets contain proprietary information. However, meaningful collaboration will still require harmonization of botanical nomenclature, metadata and analytical procedures.

14.5 AI and Precision Cultivation

AI could eventually connect environmental variables with phytochemical production.

A predictive cultivation system might integrate:

Soil properties + temperature + rainfall + irrigation + plant genotype + developmental stage → predicted phytochemical profile

This could support precision cultivation of medicinal plants and potentially improve batch-to-batch chemical consistency.

Such approaches are particularly relevant because environmental factors can substantially affect the quality and secondary-metabolite composition of medicinal plants.

14.6 Generative AI and Natural-Product-Inspired Chemistry

Generative AI may expand natural-product research from identification toward molecular design. Models can potentially generate or optimize structures inspired by naturally occurring scaffolds.

The most useful application may not be unrestricted molecular generation but constrained generation, where the system is required to satisfy biological, physicochemical, synthetic and safety criteria.

The generated compounds would then enter an experimental cycle involving synthesis, structural confirmation and biological evaluation.

14.7 Sustainable Natural-Product Discovery

AI-assisted prioritization may also contribute to conservation. Computational models could help identify the most promising plant species or tissues before large-scale collection.

This could reduce unnecessary harvesting and encourage alternative approaches such as cultivation, tissue culture, microbial production or biosynthetic reconstruction for high-value compounds.

XV. CHALLENGES AND LIMITATIONS

Although AI provides substantial opportunities, several limitations must be addressed before it can become a routine component of pharmacognosy.

Dataset Bias and Representativeness: A model trained using samples from one region may not generalize to another. Similarly, an image model trained primarily using clean laboratory photographs may fail under natural field conditions.

Therefore, dataset diversity should be regarded as a fundamental component of model validation rather than an optional enhancement.

Taxonomic Errors in Training Data: AI cannot correct systematic errors in its training labels automatically. If a reference specimen is incorrectly identified, the model may learn the wrong association.

Botanical authentication of reference datasets should therefore involve qualified taxonomic experts and, where appropriate, voucher specimens and molecular confirmation.

Analytical Batch Effects: Metabolomic and spectroscopic models may learn differences caused by analytical instruments rather than genuine biological variation. Standardized preprocessing, calibration and external validation are therefore necessary.

Model Explainability: Black-box models remain difficult to evaluate scientifically. This is particularly problematic when AI predictions are used for quality-control or regulatory decisions.

Overfitting and Data Leakage: Extremely high accuracy can sometimes indicate overfitting rather than genuine biological generalization. Independent testing and biologically meaningful dataset separation are therefore essential.

Insufficient Natural-Product Data: Natural products occupy chemically diverse and structurally complex chemical space. AI models developed mainly using conventional pharmaceutical datasets may not adequately represent this diversity. The natural-product AI field therefore requires dedicated, high-quality datasets. [44]

Regulatory Validation: AI-assisted quality-control systems must eventually satisfy requirements for reproducibility, traceability and validation. A model should have a documented development history and clearly defined acceptance criteria.
A recent comprehensive review of herbal-medicine quality control identified machine learning combined with infrared spectroscopy, mass spectrometry and NMR as a rapidly developing approach for species identification, adulteration detection and geographical-origin tracing, while emphasizing the need for standardized and data-driven quality-assurance frameworks. [45]

AI Should Not Replace Experimental Validation: The most important limitation is that computational prediction is not equivalent to experimental evidence.
An AI model may predict that a phytochemical has anti-inflammatory activity, but this does not establish cellular activity, pharmacokinetics, safety or clinical efficacy. Similarly, an AI system may classify a plant as authentic, but regulatory quality control may still require pharmacopoeial testing.
Thus, AI should remain an augmentation technology within pharmacognostic science.

XVI. RECOMMENDED STANDARDIZED WORKFLOW FOR AI-BASED PHARMACOGNOSY RESEARCH

Future research projects should use a reproducible workflow.
Step 1. Define the Scientific Question
Specify whether the objective is plant identification, authentication, adulteration detection, chemical prediction, bioactivity prediction or quality assessment.

Step 2. Establish Authenticated Reference Materials
Use taxonomically verified material and maintain appropriate voucher documentation.

Step 3. Standardize Data Collection
Document imaging conditions, plant part, geographical origin, collection date, extraction procedure and analytical platform.

Step 4. Build a Representative Dataset
Include biological, geographical, seasonal and processing variability.

Step 5. Perform Quality Control
Remove duplicated observations, incorrect labels and technically defective measurements.

Step 6. Divide Data Appropriately
Use training, validation and independent testing datasets while preventing biological data leakage.

Step 7. Develop Baseline and Advanced Models
Compare conventional statistical or machine-learning models with deep-learning methods.

Step 8. Assess Interpretability
Identify the features responsible for model predictions.

Step 9. Perform External Validation
Test the model using independent samples from another collection site, laboratory or production batch.

Step 10. Experimentally Confirm Predictions
Use microscopy, chromatography, spectroscopy, molecular biology or pharmacological experiments as appropriate.

Step 11. Establish Model Governance
Record model version, training data, preprocessing parameters, validation results and update history.

Step 12. Define Human Oversight
Ensure that final scientific or regulatory decisions remain subject to qualified expert review.

XVII. COMPARATIVE PERSPECTIVE: CONVENTIONAL VERSUS AI-AUGMENTED PHARMACOGNOSY

Table 5. Conventional versus AI-Augmented Pharmacognosy

Parameter	Conventional pharmacognosy	AI-augmented pharmacognosy
Plant identification	Expert morphological examination	Computer vision + expert confirmation

Microscopy	Manual evaluation	Automated feature recognition + manual confirmation
Chemical profiling	Chromatographic/spectroscopic analysis	Fingerprint analysis + ML
Dereplication	Manual database comparison	AI-assisted prioritization
Bioactivity screening	Predominantly experimental	Computational prioritization + experimentation
Target identification	Literature and laboratory studies	AI-assisted prediction + experimental validation
Adulteration detection	Marker analysis and expert interpretation	Multivariate fingerprint classification
Geographical authentication	Chemical analysis	ML-based chemical pattern recognition
Batch consistency	Conventional statistical assessment	Predictive multivariate quality models
Literature mining	Manual searching	NLP-assisted information extraction
Knowledge integration	Separate datasets	Multimodal AI and knowledge graphs

The comparative model indicates that AI does not eliminate conventional pharmacognosy. Instead, it creates an additional computational layer capable of processing larger and more complex datasets.

XVIII. CONCLUSION

Artificial intelligence is emerging as a transformative technology in pharmacognosy, particularly in medicinal plant identification, phytochemical discovery and herbal drug standardization. Computer vision and deep learning can support rapid recognition of medicinal plants, while machine learning applied to chromatographic, spectroscopic and metabolomic fingerprints can improve authentication, adulteration detection and quality assessment.

At the phytochemical-discovery level, AI can support dereplication, metabolite annotation, biological-activity prediction, target identification, biosynthetic-pathway analysis and molecular prioritization. Integration with genomics, transcriptomics, proteomics and metabolomics further creates the possibility of systems-level natural-product discovery.

The greatest opportunity lies in connecting these applications into a unified digital pharmacognostic workflow. Such a system could begin with an authenticated medicinal plant, integrate visual and molecular identification, characterize its chemical fingerprint, prioritize bioactive constituents and continuously evaluate the quality of the resulting herbal product.

Nevertheless, AI should not be interpreted as a replacement for pharmacognostic expertise or experimental science. Dataset quality, botanical accuracy, analytical variability, model explainability, external validation and regulatory acceptance remain substantial challenges. The future success of AI in pharmacognosy will therefore depend on high-quality reference datasets, interdisciplinary collaboration, transparent algorithms and rigorous experimental validation.

The most appropriate vision is consequently not “AI replacing pharmacognosy,” but “AI-augmented pharmacognosy,” in which computational intelligence expands the capacity of pharmacognosists to identify plants, discover phytochemicals and establish reproducible herbal-drug quality.

ACKNOWLEDGEMENT

The authors acknowledge the contributions of researchers, botanical institutions, analytical laboratories and scientific databases that support the development of computational approaches for medicinal-plant research and natural-product science.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest associated with the preparation of this review.

FUNDING STATEMENT

No specific funding was received for the preparation of this review article.

REFERENCES

[1] Mullowney, M. W., Duncan, K. R., Elsayed, S. S., Garg, N., van der Hooft, J. J. J., Martin, N. I., et al. “Artificial intelligence for natural product

drug discovery." Nature Reviews Drug Discovery, vol. 22, pp. 895–916, 2023.

[2] "Assessing deep convolutional neural network models and their comparative performance for automated medicinal plant identification from leaf images." Heliyon, vol. 10, no. 1, 2024.

[3] Newman, D. J., and Cragg, G. M. "Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019." Journal of Natural Products, vol. 83, pp. 770–803, 2020.

[4] Terlouw, B. R., Blin, K., Navarro-Muñoz, J. C., Avalon, N. E., Chevrette, M. G., Egbert, D. B., et al. "MIBiG 3.0: A community-driven effort to annotate experimentally validated biosynthetic gene clusters." Nucleic Acids Research, vol. 51, 2023.

[5] van der Hooft, J. J. J., Vervoort, J., Bino, R. J., and de Vos, R. C. H. "Spectral trees as a robust tool in chemical annotation of complex plant extracts." Analytical Chemistry, vol. 83, pp. 129–136, 2011.

[6] van der Hooft, J. J. J., Vervoort, J., Ridder, L., Jacobs, D. M., Hall, R. D., and Roger van den H. "Predicting the chemical structure of novel natural products from tandem mass spectra." Natural Product Reports, vol. 28, pp. 1801–1812, 2011.

[7] Kang, K. B., Kubátová, A., Boonphong, S., Lee, H. S., Seo, E. J., et al. "Artificial intelligence and machine learning applications in natural product research." Phytochemistry Reviews, 2024.

[8] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., et al. "Applications of machine learning in drug discovery and development." Nature Reviews Drug Discovery, vol. 18, pp. 463–477, 2019.

[9] Wäldchen, J., and Mäder, P. "Plant species identification using computer vision techniques: A systematic literature review." Archives of Computational Methods in Engineering, vol. 25, pp. 507–543, 2018.

[10] Barré, P., Stoeckle, M., Cadot, M., Piry, S., Bonhomme, V., et al. "Automated plant identification using deep learning and leaf images." Applications in Plant Sciences, 2022.

[11] Lee, S. H., Chan, C. S., Wilkin, P., and Remagnino, P. "Deep-plant: Plant identification with convolutional neural networks." IEEE International Conference on Image Processing, pp. 452–456, 2015.

[12] Mohanty, S. P., Hughes, D. P., and Salathé, M. "Using deep learning for image-based plant disease detection." Frontiers in Plant Science, vol. 7, 1419, 2016.

[13] Kaur, P., Kaur, S., and Singh, B. "Machine learning approaches in medicinal plant identification and classification: Recent developments and perspectives." Computers in Biology and Medicine, 2023.

[14] Ratnasekhar, C. H., Rai, A. K., Rakwal, P., Khan, S., Verma, A. K., Mukhopadhyay, P., et al. "Machine learning-guided Orbitrap-HRAMS-based metabolomic fingerprinting for geographical origin, variety and tissue specific authentication, and adulteration detection of turmeric and ashwagandha." Food Chemistry, vol. 482, 144078, 2025.

[15] Shen, Y., Du, H., Liu, F., Gou, X., Tao, Y., Lan, X., et al. "Quality control of herbal medicine based on analytical techniques and machine learning: Current advances and future perspectives." Phytochemical Analysis, vol. 37, no. 4, pp. 488–506, 2026.

[16] World Health Organization. Quality Control Methods for Medicinal Plant Materials. Geneva: World Health Organization, 1998.

[17] Heinrich, M., Appendino, G., Efferth, T., Fürst, R., Izzo, A. A., Kayser, O., et al. "Best practice in research—Overcoming common challenges in phytopharmacological research." Journal of Ethnopharmacology, vol. 246, 112230, 2020.

[18] Atanasov, A. G., Zotchev, S. B., Dirsch, V. M., and Supuran, C. T. "Natural products in drug discovery: Advances and opportunities." Nature Reviews Drug Discovery, vol. 20, pp. 200–216, 2021.

[19] Sarker, S. D., and Nahar, L. "Natural products isolation." Methods in Molecular Biology, vol. 864, pp. 1–25, 2012.

[20] Wolfender, J. L., Nuzillard, J. M., van der Hooft, J. J. J., Renault, J. H., and Bertrand, S. "Accelerating metabolite identification in natural product research: Toward an ideal combination of liquid chromatography-high-resolution tandem mass spectrometry profiling and automated structure annotation." Phytochemistry, vol. 161, pp. 9–22, 2019.

[21] Blaženović, I., Kind, T., Ji, J., and Fiehn, O. "Software tools and approaches for compound identification of LC-MS/MS data in metabolomics." Metabolites, vol. 8, 31, 2018.

[22] Allwood, J. W., Ellis, D. I., Heald, J. K., Heaton, D., Mur, L. A. J., Jones, M., et al. "Metabolomics and machine learning for plant research." Plant Methods, 2021.

[23] Wishart, D. S. "Metabolomics for investigating physiological and pathophysiological processes." Physiological Reviews, vol. 99, pp. 1819–1875, 2019.

[24] Dunn, W. B., Erban, A., Weber, R. J. M., Creek, D. J., Brown, M., Breitling, R., et al. "Mass appeal: Metabolite identification in mass spectrometry-focused untargeted metabolomics." Metabolomics, vol. 9, pp. 44–66, 2013.

[25] Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Grevoz, V., et al. "The seco-iridoid pathway from the medicinal plant Catharanthus roseus." Nature Communications, vol. 5, 3606, 2014.

[26] Goossens, J., et al. "Integrating multi-omics approaches for understanding plant specialized metabolism." Plant Journal, 2023.

[27] "AI-driven multi-omics approaches in natural-product drug discovery and development." 2026.

[28] Hasin, Y., Seldin, M., and Lusis, A. "Multi-omics approaches to disease." Genome Biology, vol. 18, 83, 2017.

[29] Kibble, M., Chen, J., et al. "Network pharmacology approaches for natural products and herbal medicines." Frontiers in Pharmacology, 2022.

[30] Singh, S., et al. "Deep learning approaches for medicinal plant identification: Recent developments and applications." Heliyon, 2024.

[31] Ichim, M. C. "The DNA-based authentication of commercial herbal products reveals their ingredient trade." Frontiers in Plant Science, vol. 10, 1390, 2019.

[32] Ratnasekhar, C. H., Rai, A. K., Rakwal, P., Khan, S., Verma, A. K., Mukhopadhyay, P., et al. "Machine learning-guided Orbitrap-HRAMS-based metabolomic fingerprinting for geographical origin, variety and tissue specific authentication, and adulteration detection of turmeric and ashwagandha." Food Chemistry, vol. 482, 144078, 2025.

[33] Dhami, N., and Mishra, A. D. "Phytochemical variation and quality assessment of medicinal plants: Implications for authentication." Journal of Applied Research on Medicinal and Aromatic Plants, 2022.

[34] Shen, Y., Du, H., Liu, F., Gou, X., Tao, Y., Lan, X., et al. "Quality control of herbal medicine based on analytical techniques and machine learning: Current advances and future perspectives." Phytochemical Analysis, vol. 37, no. 4, pp. 488–506, 2026.

[35] Shen, Y., Du, H., Liu, F., Gou, X., Tao, Y., Lan, X., et al. "Quality control of herbal medicine based on analytical techniques and machine learning: Current advances and future perspectives." Phytochemical Analysis, vol. 37, no. 4, pp. 488–506, 2026.

[36] Sahu, S., et al. "Deep learning for medicinal plant recognition: A systematic review." Frontiers in Plant Science, 2023.

[37] Ratnasekhar, C. H., Rai, A. K., Rakwal, P., Khan, S., Verma, A. K., Mukhopadhyay, P., et al. "Machine learning-guided Orbitrap-HRAMS-based metabolomic fingerprinting for geographical origin, variety and tissue specific authentication, and adulteration detection of turmeric and ashwagandha." Food Chemistry, vol. 482, 144078, 2025.

[38] Mullowney, M. W., Duncan, K. R., Elsayed, S. S., Garg, N., van der Hooft, J. J. J., Martin, N. I., et al. "Artificial intelligence for natural product drug discovery." Nature Reviews Drug Discovery, vol. 22, pp. 895–916, 2023.

[39] "Artificial intelligence and multi-omics in natural-product drug discovery and development." 2026.

[40] Mullowney, M. W., Duncan, K. R., Elsayed, S. S., Garg, N., van der Hooft, J. J. J., Martin, N. I., et al. "Artificial intelligence for natural product drug discovery." Nature Reviews Drug Discovery, vol. 22, pp. 895–916, 2023.

[41] Shen, Y., Du, H., Liu, F., Gou, X., Tao, Y., Lan, X., et al. "Quality control of herbal medicine based on analytical techniques and machine learning: Current advances and future perspectives." Phytochemical Analysis, vol. 37, no. 4, pp. 488–506, 2026.

[42] Shen, Y., Du, H., Liu, F., Gou, X., Tao, Y., Lan, X., et al. “Quality control of herbal medicine based on analytical techniques and machine learning: Current advances and future perspectives.” Phytochemical Analysis, vol. 37, no. 4, pp. 488–506, 2026.

[43] Mullowney, M. W., Duncan, K. R., Elsayed, S. S., Garg, N., van der Hooft, J. J. J., Martin, N. I., et al. “Artificial intelligence for natural product drug discovery.” Nature Reviews Drug Discovery, vol. 22, pp. 895–916, 2023.

[44] Mullowney, M. W., Duncan, K. R., Elsayed, S. S., Garg, N., van der Hooft, J. J. J., Martin, N. I., et al. “Artificial intelligence for natural product drug discovery.” Nature Reviews Drug Discovery, vol. 22, pp. 895–916, 2023.

[45] Shen, Y., Du, H., Liu, F., Gou, X., Tao, Y., Lan, X., et al. “Quality control of herbal medicine based on analytical techniques and machine learning: Current advances and future perspectives.” Phytochemical Analysis, vol. 37, no. 4, pp. 488–506, 2026.