INTERNATIONAL JOURNAL OF PHARMACEUTICAL SCIENCES
[ISSN: 0975-4725; CODEN(USA): IJPS00]
Journal Homepage: https://www.ijpsjournal.com

OPEN ACCESS

Review Paper

Artificial Intelligence-Assisted Herbal Formulation Development: From Phytochemical Intelligence to Predictive Nanodelivery and Precision Phytotherapy

Alisha Jabi[1]*, Deepika Shakya[2], Priti Yadav[3]

[1,3]Lecturer, Smt. Vidyawati College of Pharmacy, Goramachhiya Jhansi
[2] Phd Scholar, Institute of Pharmacy, Bundelkhand University, Jhansi.

ARTICLE INFO

Published: 21 Sept 2026
Keywords:
Artificial intelligence; machine learning; herbal formulation; natural products; phytopharmaceuticals; phytochemical profiling; nanocarriers; formulation optimisation; Quality by Design; precision phytotherapy
DOI:
10.5281/zenodo.22868186

ABSTRACT

Herbal medicines constitute a chemically diverse source of bioactive molecules and remain important components of traditional and complementary healthcare systems. However, the development of reproducible pharmaceutical formulations from herbal materials is complicated by variability in botanical identity, geographical origin, cultivation, harvesting, processing, extraction, phytochemical composition, and storage, all of which contribute to batch-to-batch variation in biological activity and formulation performance. The presence of multiple co-occurring constituents further complicates the prediction of pharmacological activity, stability, bioavailability, and therapeutic response, while conventional herbal formulation development remains largely dependent on empirical, trial-and-error experimentation that becomes inefficient as the number of interacting formulation and process variables increases.Artificial intelligence (AI) — encompassing machine learning, deep learning, artificial neural networks, computer vision, natural language processing, quantitative structure–activity/property relationship modelling, graph neural networks, and generative approaches — offers an emerging, data-driven framework for addressing these limitations. This review critically examines AI applications spanning the herbal formulation-development continuum, including medicinal-plant authentication, phytochemical fingerprint interpretation, bioactive-compound prioritisation, herb–compound–target and network pharmacology analysis, polyherbal synergy prediction, nanocarrier selection and optimisation, Quality-by-Design and critical-quality-attribute prediction, dissolution and release modelling, stability prediction, quality control and adulteration detection, ADME and toxicity screening, and personalised phytopharmaceutical development. Particular attention is given to the convergence of AI with nanotechnology, given the poor aqueous solubility, limited permeability, and instability characteristic of many phytoconstituents.

***Corresponding Author:** Alisha Jabi
Address: *Lecturer, Smt. Vidyawati College of Pharmacy, Goramachhiya Jhansi.*
Email ✉: alishamansuri141@gmail.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Despite this progress, translation into routine practice remains constrained by small and heterogeneous datasets, herbal batch variability, inconsistent metadata reporting, limited external validation, domain shift, model interpretability, and evolving regulatory frameworks. Future progress will depend on standardised datasets, explainable and physics-informed AI, automated closed-loop experimentation, multimodal data integration, digital twins, and prospective experimental validation. AI should therefore be regarded as a decision-support technology that complements, rather than replaces, pharmaceutical experimentation in herbal formulation science

INTRODUCTION

Medicinal plants have long constituted a principal source of therapeutic agents and continue to occupy a central position in natural-product research. Owing to their remarkable structural and chemical diversity, natural products have furnished molecular scaffolds for a wide range of pharmacologically significant compounds. Despite this contribution, the translation of traditional herbal knowledge into standardized pharmaceutical products remains a persistent challenge, largely because herbal preparations constitute multicomponent systems rather than discrete molecular entities [1–4].

The chemical composition of herbal material is governed by a multiplicity of factors, including botanical species, genotype, geographical origin, soil characteristics, climatic conditions, cultivation practices, harvest timing, and post-harvest handling and storage. As a result, considerable batch-to-batch variability may arise even among materials derived from ostensibly identical botanical sources. Such variability, in turn, has direct implications for biological activity, pharmacokinetic behaviour, physicochemical stability, and formulation performance [2, 5].

This inherent complexity is further compounded when herbal extracts are incorporated into pharmaceutical dosage forms. Individual phytoconstituents may vary considerably in molecular weight, polarity, lipophilicity, aqueous solubility, membrane permeability, and chemical stability. Certain compounds are prone to rapid metabolic clearance or limited gastrointestinal absorption, while others may degrade during manufacturing or storage. Effective herbal formulation development therefore necessitates the concurrent evaluation of botanical quality, phytochemical composition, formulation-related variables, and biological performance.

Artificial intelligence (AI) has emerged as a promising strategy for addressing datasets of this multidimensional nature. Machine learning (ML) and deep learning (DL) approaches are capable of discerning nonlinear relationships between independent variables and experimental outcomes, thereby facilitating prediction, classification, and optimization tasks. Recent reviews have underscored the expanding role of AI across natural-product discovery, pharmacological prediction, target identification, formulation design, and drug-delivery system development [3, 6–9].

The convergence of AI and herbal medicine holds particular promise given that traditional herbal research generates highly heterogeneous data spanning phytochemistry, pharmacognosy, analytical chemistry, pharmacology, toxicology, and pharmaceutical technology. AI-based approaches offer the potential to consolidate these disparate data streams within a unified computational framework. Notably, recent work has described AI-assisted integration of omics data, network pharmacology, and herbal medicine research, while other investigations have explored AI-driven prediction of personalised nanocarrier formulations for herbal therapeutics [10, 11].

2. Artificial Intelligence Techniques Applicable to Herbal Formulation Development

Artificial intelligence does not constitute a single, monolithic computational method; rather, it encompasses a diverse spectrum of algorithmic

approaches, each characterised by distinct underlying architectures, data requirements, and domains of applicability. Several of these approaches bear direct and substantial relevance to the multidimensional challenges inherent in herbal formulation development, ranging from raw-material characterisation to formulation optimisation and quality assurance.

2.1 Machine Learning–Based Predictive Modelling

Machine learning algorithms are designed to identify underlying, often nonlinear, relationships between input variables and experimentally observed outcomes, thereby enabling predictive modelling in the absence of an explicit mechanistic equation. Approaches commonly employed in formulation-related research include Random Forest, Support Vector Machine, Support Vector Regression, Artificial Neural Networks, Gradient Boosting, XGBoost, k-Nearest Neighbour, Partial Least Squares Regression, and Gaussian Process Regression. Each of these algorithms differs in its underlying assumptions, computational cost, and suitability for datasets of varying size and dimensionality, and the selection of an appropriate algorithm is therefore contingent upon the specific formulation problem under investigation.

Within the context of formulation research, independent variables typically encompass the concentrations of extract, polymer, lipid, surfactant, or co-surfactant, in addition to relevant manufacturing parameters such as homogenisation speed, sonication time, or temperature. Corresponding dependent variables may include particle size, polydispersity index, zeta potential, encapsulation efficiency, drug loading, and in vitro release profile. Evidence from recent formulation studies indicates that ML-based approaches can effectively navigate multidimensional formulation spaces that would otherwise be difficult to explore comprehensively through conventional trial-and-error experimentation or classical design-of-experiments methodology, thereby reducing the number of experimental runs required to identify an optimised formulation [7, 12].

2.2 Deep Learning Architectures for Complex Pattern Recognition

Deep learning methods are capable of extracting complex, nonlinear, and hierarchical patterns from large-scale datasets without extensive manual feature engineering. Convolutional neural networks (CNNs) are particularly well suited to image-based applications such as plant authentication and histological examination, owing to their capacity to automatically learn spatial features from raw image data. Graph-based architectures, by contrast, offer distinct advantages for modelling molecular structures and drug–target relationships, as they can inherently represent the connectivity and topology of chemical entities.

2.3 Artificial Neural Networks in Formulation Optimisation

Artificial neural networks (ANNs) enable the modelling of nonlinear relationships between formulation variables and critical quality attributes, functioning as universal function approximators capable of capturing interactions that are not readily described by conventional polynomial models. Their utility is especially apparent in situations where experimental responses are governed by multiple interacting factors, a circumstance frequently encountered in herbal formulation systems owing to the multicomponent nature of plant extracts. ANN-based models are frequently combined with optimisation algorithms, such as genetic algorithms or particle swarm optimisation, to identify formulation compositions that satisfy

multiple, sometimes competing, quality targets simultaneously.

2.4 Natural Language Processing for Knowledge Extraction

Natural language processing (NLP) techniques can be applied to the systematic analysis of scientific publications, traditional medicine records, patent literature, and ethnopharmacological sources. Such approaches offer a means of converting unstructured traditional knowledge — often documented in classical texts or regional languages — into structured, machine-readable datasets amenable to subsequent computational analysis. This capability is particularly valuable for bridging traditional ethnobotanical knowledge with contemporary evidence-based formulation science [1, 3].

2.5 Computer Vision for Botanical Authentication and Quality Assessment

Computer vision technologies, frequently underpinned by convolutional neural network architectures, have found application across several domains relevant to herbal medicine, including:

- Plant species identification and taxonomic classification
- Leaf and macroscopic morphological recognition
- Powdered drug and raw-material authentication
- Microscopic examination of histological features
- Detection of adulteration or substitution with inferior or spurious material
- Automated quality classification and grading

Collectively, these applications address a longstanding challenge in herbal medicine, namely the reliable and reproducible authentication of botanical raw material prior to formulation.

2.6 Graph Neural Networks for Molecular and Network-Level Analysis

Graph neural networks (GNNs) represent molecular structures and biological relationships in graph form, wherein atoms or biological entities constitute nodes and their interactions constitute edges. This representation enables their application to compound–target prediction, molecular similarity analysis, and the identification of potential synergistic interactions among phytoconstituents — an area of particular relevance to herbal medicine, given that therapeutic effects are frequently attributed to the combined action of multiple constituents rather than a single active molecule [3,10].

2.7 Generative Artificial Intelligence in Molecular and Formulation Design

Generative models offer potential utility in the generation of novel molecular structures, the prediction of structural modifications, and the design of candidate compounds or formulation compositions that satisfy predefined physicochemical or biological criteria. It should be emphasised, however, that computationally generated candidates remain hypothetical in nature and require rigorous experimental validation before any conclusions regarding their pharmacological or formulation-related utility can be drawn.

Table 1. AI technologies and their applications in herbal formulation development

AI approach	Herbal-pharmaceutical application	Major output
Machine learning	Formulation optimization	Optimized formulation
Deep learning	Complex chemical datasets	Pattern recognition
ANN	Nonlinear formulation prediction	CQA prediction
Computer vision	Plant authentication	Species identification
NLP	Traditional medicine literature	Knowledge extraction
QSAR/QSPR	Phytochemical prediction	Activity/property prediction
GNN	Herb–compound–target analysis	Target prediction
Generative AI	Molecular/formulation design	Candidate generation
Explainable AI	Model interpretation	Feature importance
Physics-informed ML	Release/stability/toxicity	Mechanistically constrained prediction

3 Artificial Intelligence in Herbal Raw-Material Authentication

Authentication constitutes the first critical step in herbal formulation development, as the incorrect identification or inadvertent substitution of plant material may compromise both therapeutic efficacy and patient safety. Given that herbal raw materials are frequently sourced from diverse geographical regions and supply chains, the risk of misidentification, adulteration, or intentional substitution with morphologically similar but pharmacologically inferior species remains a persistent concern in the herbal medicine sector.

Traditional authentication approaches commonly rely on the following:

- Morphological examination
- Microscopic evaluation
- Macroscopic characterisation
- Physicochemical assessment
- Chromatographic analysis
- Spectroscopic analysis
- DNA-based molecular approaches

While these approaches remain indispensable and continue to form the foundation of pharmacognostic practice, they present certain practical limitations. Several require specialised taxonomic or analytical expertise that may not be uniformly available, and the associated procedures can be comparatively time-consuming when large numbers of samples require processing, as is often the case in industrial-scale herbal material procurement.

Artificial intelligence offers an opportunity to reframe authentication as a pattern-recognition problem, wherein spectroscopic or chromatographic fingerprints are converted into numerical datasets and subsequently subjected to classification algorithms capable of discriminating between species, varieties, or geographical origins with a high degree of accuracy and reproducibility. In this regard, recent work has demonstrated the integration of attenuated total reflectance Fourier-transform infrared (ATR-FTIR) fingerprints with multiple chemometric and machine-learning methods for the high-throughput discrimination of fifty-three root and rhizome Chinese herbal materials. The optimised support vector machine approach reported in this study achieved a classification accuracy of 100% across both the training and validation datasets [13]. In a related investigation, machine-learning-guided Orbitrap high-resolution mass spectrometry metabolomic fingerprinting was employed for the authentication of turmeric and ashwagandha according to geographical origin, botanical variety, and plant tissue, with the approach additionally proving capable of detecting adulteration [14].

Collectively, these findings suggest that AI-based methods are best positioned to complement, rather than supplant, conventional pharmacognostic authentication — offering enhanced throughput and discriminatory power while remaining dependent on the analytical and reference standards established by classical approaches.

4. Artificial Intelligence in Phytochemical Profiling

Herbal extracts contain numerous chemically diverse constituents, the composition of which may vary considerably depending on botanical, environmental, and processing factors discussed previously. Conventional analytical approaches frequently focus on the quantification of one or a small number of selected marker compounds; however, such single-marker analysis may not adequately capture the full chemical complexity of a given botanical preparation, and may therefore provide an incomplete basis for quality assessment or biological correlation.

Modern analytical platforms — including high-performance liquid chromatography (HPLC), liquid chromatography–mass spectrometry (LC-MS), gas chromatography–mass spectrometry (GC-MS), nuclear magnetic resonance (NMR) spectroscopy, Fourier-transform infrared (FTIR) spectroscopy, Raman spectroscopy, and near-infrared (NIR) spectroscopy — generate high-dimensional datasets that are well suited to interpretation using machine-learning-based approaches, given the difficulty of extracting meaningful patterns from such data through manual inspection alone.

Within this context, artificial intelligence can contribute to phytochemical profiling in several respects, including:

1. Classification of herbal extracts according to species, variety, or quality grade
2. Identification of characteristic chemical fingerprints associated with specific botanical sources
3. Detection of batch-to-batch chemical variability
4. Recognition of geographically determined chemical differences
5. Identification of potential marker compounds
6. Prediction of biological activity from chemical composition
7. Detection of adulteration or substitution

Illustrating this potential, recent research employing FT-NIR spectroscopy in conjunction with electronic-eye and electronic-nose technologies, combined with machine-learning analysis, demonstrated highly accurate origin identification and quality assessment of *Curcuma* samples, with reported prediction performance exceeding 0.99 for selected constituents [15].

Taken together, these developments indicate that AI-assisted phytochemical profiling offers a means of transitioning from a single-marker analytical paradigm toward a more comprehensive, multivariate approach to chemical characterisation — one that is arguably better aligned with the inherently multicomponent nature of herbal preparations.

5. AI in Bioactive Compound Identification

The identification of biologically active compounds represents one of the most critical, and historically most resource-intensive, stages of natural-product research. Traditional workflows for bioactive compound discovery typically proceed through a sequential pathway of extraction, fractionation, isolation, structural elucidation, and biological testing. Although this approach has yielded substantial scientific value over several decades and remains the definitive means of confirming bioactivity, it is frequently characterised by considerable demands on time, material resources, and analytical infrastructure,

particularly when applied to complex botanical matrices containing numerous co-occurring constituents.

Artificial intelligence offers the possibility of introducing a computational prioritisation stage prior to resource-intensive experimental work, thereby reorganising the discovery pathway along the following lines: phytochemical database compilation, molecular representation, AI-based prediction, candidate ranking, and, finally, experimental validation of the most promising candidates. Within this framework, machine-learning models can be employed to estimate a range of parameters relevant to compound prioritisation, including:

- Biological activity
- Target binding affinity
- Structural similarity to known active compounds
- Absorption, distribution, metabolism, and excretion (ADME) properties
- Toxicological risk
- Drug-likeness

Supporting the growing adoption of this approach, a recent review reported that AI is increasingly being applied to screen natural-product datasets for pharmacological activity and disease-associated molecular targets [6]. In a related account, tree-based models, graph neural networks, molecular embeddings, network pharmacology, and multi-omics integration have been described as emerging components of AI-assisted natural-product discovery, reflecting the methodological diversity now being brought to bear on this problem [3]. Consistent with these observations, additional literature has similarly characterised AI as a computational tool with applicability to natural-product molecular design and lead discovery more broadly [16].

6. AI in Herb–Compound–Target Analysis

The pharmacological behaviour of herbal medicines frequently arises not from the action of a single compound at a single molecular target, but rather from complex interactions among multiple phytochemicals acting concurrently upon multiple molecular targets. This multi-component, multi-target characteristic can be conceptually represented as a sequential relationship:

Herb → Phytochemicals → Molecular targets → Biological pathways → Overall biological response.

Network pharmacology provides a useful conceptual and analytical framework for examining these relationships in a systems-level context, while artificial intelligence contributes the computational capacity required to improve prediction accuracy and to integrate the large-scale, heterogeneous biological datasets upon which such analyses depend. In this regard, recent work has specifically examined the application of AI-powered omics integration and network pharmacology to the study of complex herbal medicines [10]. This approach combines AI-based analytical methods with genomic, transcriptomic, proteomic, and metabolomic data in order to investigate the underlying biological mechanisms through which herbal preparations exert their effects.

By integrating these diverse data streams, AI-assisted network pharmacology approaches may assist in distinguishing among the following:

- Primary active compounds
- Secondary or minor contributing constituents
- Potential synergistic interactions among phytoconstituents
- Potential antagonistic interactions among phytoconstituents
- Disease-associated biological pathways
- Candidate therapeutic targets warranting further investigation

It must be emphasised, however, that interactions identified through computational network

pharmacology approaches represent predictions derived from statistical and topological association, and should not be interpreted as confirmed pharmacological mechanisms in the absence of corresponding experimental validation. Computational predictions of this nature are most appropriately regarded as hypothesis-generating tools that inform, rather than replace, subsequent mechanistic and pharmacological investigation.

7. AI in Polyherbal Formulation

Polyherbal formulations present a particularly challenging case within herbal formulation development, as several plant extracts may be combined at varying ratios to achieve a desired therapeutic outcome. For instance, the combination of Extract A, Extract B, and Extract C may generate a pharmacological response that cannot be reliably predicted by considering the effects of each constituent extract in isolation, owing to the potential for complex interactions among their respective phytoconstituents.

Within this context, artificial intelligence offers the potential to model several interrelated aspects of polyherbal formulation, including:

- Extract-to-extract ratios
- Marker-compound concentrations
- Resultant biological activity
- Synergistic interactions among constituents
- Potential toxicological interactions
- Formulation stability
- Overall formulation performance

Consistent with this potential, recent literature addressing the intersection of AI and natural products has emphasised network-based modelling of herb–ingredient–target–pathway relationships, together with the computational prediction of synergistic effects among combined botanical constituents [3,10].

An AI-assisted strategy for polyherbal formulation development may accordingly proceed along the following sequence: identification of candidate herbal materials, phytochemical characterisation, compilation of relevant pharmacological datasets, AI-based prediction of synergistic interactions, determination of an optimal herb-to-herb ratio, experimental validation of the predicted formulation, and, finally, development of the finalised polyherbal preparation.

8. AI in Herbal Nanocarriers

Poor bioavailability represents a major limitation affecting numerous phytoconstituents, frequently arising from low aqueous solubility, limited membrane permeability, extensive first-pass metabolism, or physicochemical instability under physiological conditions. Nanocarrier-based delivery systems have accordingly attracted considerable attention as a means of potentially improving the solubility, stability, absorption, and tissue distribution of such constituents.

Nanocarrier platforms of particular relevance to herbal drug delivery include the following:

- Liposomes
- Phytosomes
- Nanoemulsions
- Solid lipid nanoparticles
- Nanostructured lipid carriers
- Polymeric nanoparticles
- Niosomes
- Nanocrystals
- Polymeric micelles

Given the considerable number of formulation variables associated with each of these platforms — including carrier composition, particle size, surface charge, and manufacturing method — the selection and optimisation of an appropriate nanocarrier system for a given phytoconstituent constitutes a complex, multidimensional problem well suited to AI-based approaches. In this regard, a recent review specifically examined AI-based prediction of personalised nanocarrier formulations for herbal drugs, identifying machine learning, deep learning, transformer models, and

graph neural networks as promising computational approaches for nanocarrier selection and optimisation [11].

8.1 Liposomes

Liposomes are capable of accommodating both hydrophilic and lipophilic substances within their aqueous core and lipid bilayer, respectively, and may thereby improve the stability and pharmacokinetic behaviour of encapsulated herbal compounds. A recent review addressing liposomal delivery of natural herbal constituents highlighted the potential of microfluidic fabrication approaches for achieving precise control over particle size and encapsulation efficiency, parameters that are otherwise difficult to regulate consistently using conventional preparation methods [17].

8.2 Phytosomes

Phytosomes are of particular relevance to herbal formulation development, as the complexation of phytoconstituents with phospholipids may improve their absorption and overall bioavailability relative to the unformulated extract. In this context, a recent review concluded that phytosomal delivery systems can address several longstanding limitations associated with herbal products, including inadequate solubility, poor absorption, limited bioavailability, and insufficient stability [18].

8.3 Nanoemulsions

Nanoemulsions can enhance the apparent solubility and dispersion of poorly water-soluble phytoconstituents through the formation of thermodynamically or kinetically stable oil-in-water or water-in-oil systems, and are potentially applicable across a range of administration routes, including oral and topical delivery, among others.

1.8.4 Lipid and Polymeric Nanoparticles

Lipid and polymeric nanoparticles are capable of providing controlled or sustained release profiles, together with a measure of protection against the chemical or enzymatic degradation of encapsulated phytoconstituents.

9. AI-Based Formulation Optimization

Formulation optimisation is arguably the most direct and readily applicable domain of AI implementation within pharmaceutical technology, given that formulation development inherently involves the systematic manipulation of numerous interacting variables in pursuit of a defined set of quality targets. A typical herbal formulation may be characterised by variables spanning three broad categories.

Material variables commonly include extract concentration, polymer concentration, lipid concentration, surfactant concentration, co-surfactant concentration, and phospholipid concentration.

Process variables commonly include temperature, mixing speed, homogenisation time, sonication time, solvent ratio, and drying conditions.

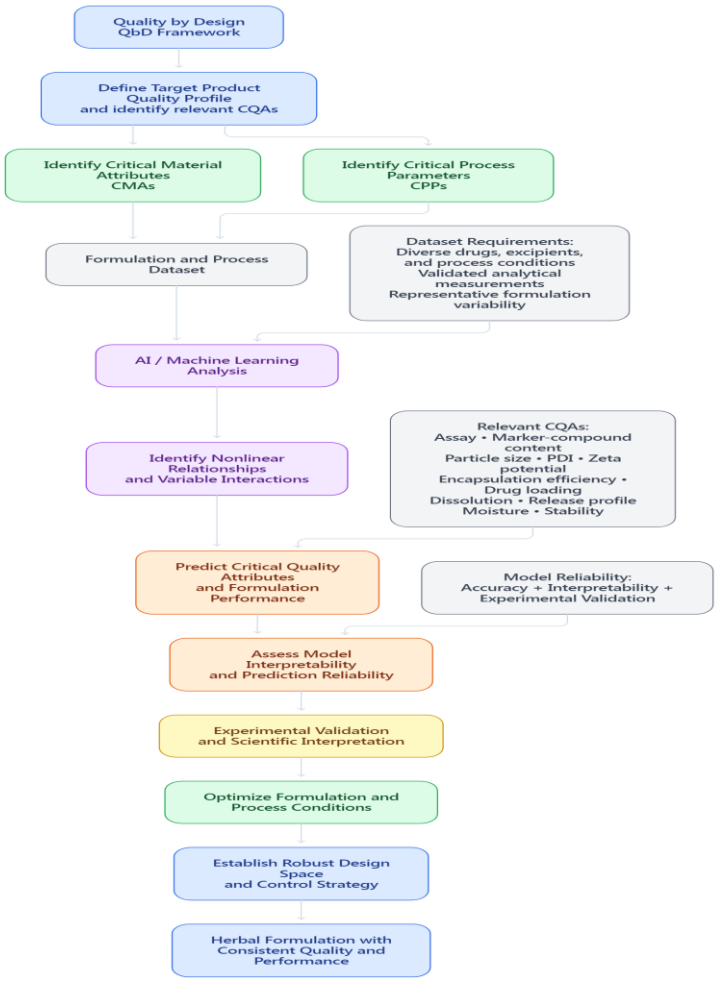
Response variables, representing the critical quality attributes against which formulation performance is ultimately assessed, commonly include particle size, polydispersity index (PDI), zeta potential, encapsulation efficiency, drug loading, dissolution behaviour, release profile, and stability.

The sheer number and interdependence of these variables render exhaustive experimental exploration impractical using conventional approaches alone, a limitation that has prompted increasing interest in machine-learning-based optimisation strategies. In this regard, recent research has demonstrated the broad applicability of machine learning across a wide range of

nanoparticulate delivery systems, including polymeric nanoparticles, lipid nanoparticles, liposomes, solid lipid nanoparticles, nanostructured lipid carriers, nanoemulsions, niosomes, and nanocrystals [12].

The general computational framework underlying this approach may be summarised as follows: formulation variables are first compiled into a structured dataset, which is then used to train a machine-learning model capable of predicting the resulting formulation responses; multi-objective optimisation algorithms are subsequently applied to identify formulation compositions that best satisfy the desired combination of quality attributes; and the resulting candidate formulations are finally subjected to experimental confirmation. By enabling the computational prediction of formulation outcomes prior to laboratory preparation, this approach offers the potential to substantially reduce the number of experimental combinations that must be physically prepared and tested, thereby improving the efficiency of the overall formulation development process while conserving material and analytical resources.

10. AI in Quality by Design and Critical Quality Attributes



11. AI in Dissolution and Release

Dissolution and release testing are critical for determining pharmaceutical performance.

Herbal products present additional challenges because they may contain several constituents with different:

- Solubilities
- Diffusion coefficients

- Molecular weights
- Degradation rates
- Absorption characteristics

AI can analyse dissolution curves and predict:

- Release rate
- Release mechanism
- Time-dependent concentration
- Formulation-dependent release
- Potential in-vitro/in-vivo relationships

A recent Bentham Science review specifically examined AI-assisted dissolution profiling of herbal therapeutics and identified ML and DL as promising approaches for analysing complex herbal dissolution datasets, improving dissolution conditions and potentially strengthening in-vitro/in-vivo correlations [20].

Independent pharmaceutical research has also demonstrated that ML can predict entire dissolution profiles from formulation composition [21].

These findings suggest that AI may become useful not only for formulation selection but also for **release-profile prediction**.

12. Artificial Intelligence in Stability Prediction

Stability represents a major concern in herbal formulation development, as phytoconstituents are susceptible to a range of degradation pathways, including oxidation, hydrolysis, photodegradation, and thermal degradation. Artificial intelligence offers the capacity to analyse complex relationships between environmental and formulation-related variables — such as temperature, humidity, storage duration, and formulation composition — and their resultant effects, including chemical degradation, physical instability, and loss of biological activity.

Potential applications of AI within this domain include the following:

- Prediction of degradation rate
- Identification of critical factors governing stability
- Comparative assessment of alternative formulation compositions
- Prediction of particle growth during storage
- Prediction of marker-compound loss over time

It should be emphasised, however, that AI-derived stability predictions are most appropriately regarded as supportive evidence rather than definitive findings. Such predictions cannot substitute for the formal accelerated and long-term stability studies mandated for pharmaceutical development and regulatory approval.

13. Artificial Intelligence in Quality Control and Adulteration Detection

Quality control represents one of the most promising areas for AI application within herbal medicine. Conventional quality-control strategies typically rely on identity testing, ash value determination, extractive value assessment, moisture content analysis, marker-compound quantification, and chromatographic fingerprinting. However, herbal matrices contain numerous constituents that collectively contribute to overall product quality, a complexity that single-parameter or marker-based testing may not fully capture.

Artificial intelligence enables the analysis of entire chemical or spectral fingerprints, facilitating the classification of samples into categories such as authentic, adulterated, substituted, or substandard. Supporting this application, recent studies have demonstrated machine-learning-assisted authentication using ATR-FTIR and metabolomic fingerprinting approaches [13,14], while a further review has similarly highlighted the potential of AI for herbal authentication, adulterant detection, and standardisation more broadly [23]. The principal advantage of this approach lies in the possibility of developing rapid, high-throughput, and inherently multivariate quality-control systems, better suited to the compositional

complexity of herbal materials than conventional single-parameter testing.

14. Artificial Intelligence in ADME and Toxicity Prediction

Herbal products should not be presumed inherently safe on the basis of their botanical origin alone. Potential safety concerns associated with herbal constituents include hepatotoxicity, nephrotoxicity, cardiotoxicity, genotoxicity, herb–drug interactions, cytochrome P450 (CYP) enzyme-mediated interactions, altered drug transport, and immunological effects.

Artificial intelligence can assist in predicting absorption, distribution, metabolism, excretion (ADME), and toxicity characteristics during the early stages of development, following a general workflow in which chemical structure data are input into an AI- or ADMET-based predictive model to generate predictions of absorption, metabolism, and toxicity, which in turn inform candidate prioritisation. Recent literature has identified AI-based prediction of pharmacological properties and biological activity as an increasingly important component of natural-product development [6]. The principal benefit of this approach lies in enabling early-stage risk prioritisation; however, computational toxicity predictions require experimental confirmation before being relied upon for safety-related decision-making.

15. Artificial Intelligence and Personalised Herbal Medicine

Personalised herbal medicine represents an emerging, though as yet developing, research direction. Patient-specific responses to herbal preparations may be influenced by a range of factors, including genetic variation, age, sex, disease phenotype, metabolic status, gut microbiome composition, concomitant medication use, and environmental exposure.

Artificial intelligence offers the potential to integrate genomic, transcriptomic, proteomic, metabolomic, microbiome, and clinical data in order to predict patient-specific responses to herbal treatment. In this regard, recent research has specifically proposed integrating genomics, microbiome, and metabolomics data with AI for the development of personalised herbal nanocarrier systems [11], while related literature has similarly discussed AI-assisted personalised nanomedicine, encompassing material selection, formulation design, toxicity prediction, and patient outcome assessment [23]. At the present stage of development, personalised herbal formulation should be regarded as an emerging research concept rather than an established component of routine clinical practice.

16. Generative AI, Explainable AI, and Physics-Informed Machine Learning

16.1 Generative Artificial Intelligence

Generative AI approaches offer potential support for molecular generation, structural optimisation, natural-product analogue design, formulation hypothesis generation, literature mining, and experimental design. Recent research addressing natural products has described generative approaches as emerging tools for natural-product optimisation and for the generation of pseudo-natural compounds [24]. Nonetheless, structures generated through such approaches require computational filtering and subsequent experimental validation before any conclusions regarding their utility can be drawn.

16.2 Explainable Artificial Intelligence

High-performing AI models are sometimes characterised as operating in a "black-box"

manner, whereby the internal basis for a given prediction is not readily interpretable. This characteristic is particularly problematic within pharmaceutical development, where researchers require a clear understanding of why a given model recommends a particular formulation composition before acting upon that recommendation. Explainable AI approaches can address this limitation by identifying the relative contribution of individual formulation variables to a given prediction, employing methods such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-agnostic Explanations (LIME), feature-importance analysis, and partial dependence analysis. Recent literature has emphasised explainability as an important requirement for improving confidence in, and facilitating scientific interpretation of, AI-based drug research findings [25].

16.3 Physics-Informed Machine Learning

Physics-informed machine learning integrates experimental data with established physical or mechanistic principles, thereby constraining model predictions to remain consistent with known physicochemical behaviour. This approach is particularly relevant to phenomena such as drug diffusion, drug release, colloidal stability, particle formation, mass transport, and nanocarrier toxicity. Illustrating this application, a recent study addressing curcumin-loaded nanocarriers developed a physics-informed machine-learning framework incorporating Derjaguin–Landau–Verwey–Overbeek (DLVO) stability theory in conjunction with drug-release kinetics, demonstrating how the incorporation of mechanistic constraints could improve the plausibility and interpretability of model predictions [26]. This concept holds particular relevance to herbal formulation research, given that herbal nanoformulations frequently involve comparatively limited datasets combined with complex underlying physicochemical interactions — circumstances under which the incorporation of mechanistic knowledge may partially compensate for data scarcity.

Table 2. Herbal formulation problems and AI-based solutions

Problem	Conventional limitation	AI-assisted approach
Plant misidentification	Time-consuming expert analysis	Computer vision/ML
Adulteration	Multiple analytical tests	Spectral ML
Batch variability	Single-marker analysis	Chemical fingerprint modelling
Poor solubility	Trial-and-error carrier selection	Predictive nanocarrier selection
Formulation optimization	Large experimental design space	ML optimization
Release prediction	Multiple empirical models	ML/DL
Stability	Long experimental studies	Predictive modelling
Toxicity	Late-stage testing	AI-assisted early screening

17 Research Gaps

Despite the considerable progress outlined in the preceding sections, several limitations continue to impede the immediate and reliable translation of AI-assisted approaches into routine herbal formulation development.

17.1 Limited Dataset Size

Many published formulation studies are based on relatively small experimental datasets, often constrained by the practical limits of laboratory-scale formulation work. Deep-learning models, in particular, generally require substantially larger and more diverse datasets than are typically

available in this field in order to achieve robust and generalisable performance.

17.2 Herbal Batch Variability

The inherent natural variability of botanical raw materials introduces a source of dataset noise that is not encountered to the same extent in datasets derived from synthetic drug substances, thereby complicating model training and potentially reducing predictive reliability.

17.3 Lack of Standardised Metadata

Important contextual information — including botanical authentication details, geographical origin, extraction conditions, and phytochemical fingerprints — is not consistently reported across published studies, limiting the extent to which datasets from different sources can be meaningfully compared or combined.

17.4 Limited External Validation

A model developed using data generated within a single laboratory may not necessarily perform with equivalent accuracy when applied to samples or formulations generated under different laboratory conditions, raising concerns regarding generalisability.

17.5 Domain Shift

Variations in equipment, raw-material sourcing, formulation methodology, and analytical procedures across different research settings can substantially alter model performance, a phenomenon commonly referred to as domain shift.

17.6 Model Overfitting

A model may exhibit high predictive performance when evaluated against its training data while demonstrating comparatively poor predictive ability when applied to previously unseen formulations, reflecting a failure to generalise beyond the specific dataset on which it was developed.

17.7 Lack of Interpretability

Predictions generated by complex, black-box models can be difficult to justify on scientific or mechanistic grounds, posing a challenge for their acceptance within a field that traditionally places considerable emphasis on mechanistic understanding.

17.8 Regulatory Uncertainty

Regulatory frameworks governing the use of AI-assisted approaches in formulation-related decision-making remain in a relatively early stage of development, contributing to uncertainty regarding the evidentiary weight that may ultimately be assigned to AI-derived predictions within a formal regulatory submission.

Consistent with these observations, recent literature has identified small dataset size, data imbalance, incomplete data provenance, domain shift, and limited model interpretability as major barriers to the translation of AI into natural-product research more broadly [3,4].

FUTURE PERSPECTIVES

The future development of AI-assisted herbal formulation should move beyond isolated prediction models toward an integrated **AI–experiment–validation ecosystem**.

A proposed future workflow is:

Medicinal plant
↓
Botanical authentication
↓
Chemical fingerprinting
↓
Multi-omics integration
↓
AI-based compound prioritization

↓
Herb–compound–target analysis
↓
Synergy prediction
↓
Delivery-system selection
↓
AI-based formulation optimization
↓
CQA prediction
↓
Automated formulation preparation
↓
Analytical characterization
↓
AI model updating
↓
Pharmacokinetics and toxicity
↓
Clinical validation
↓
Personalized phytopharmaceutical product

CONCLUSION

Artificial intelligence is emerging as a powerful enabling technology for modernising herbal formulation development, with applications spanning raw-material authentication, phytochemical profiling, bioactive-compound prioritisation, polyherbal optimisation, nanocarrier selection, formulation design, quality control, stability prediction, and personalised medicine. Recent literature demonstrates rapid progress across these domains, including AI-assisted natural-product research, multi-omics integration, and predictive nanocarrier development [3,6,10–12,16,20,22,23,27,28].

Nevertheless, AI should not be regarded as a substitute for rigorous pharmaceutical experimentation — the reliability of its predictions depends on dataset quality, appropriate model selection, external validation, and biological confirmation. The most promising path forward is therefore an iterative human–AI–experiment loop, in which computational models generate hypotheses, scientists design targeted experiments, and the resulting evidence refines the model. The convergence of AI with nanotechnology, multi-omics, quality-by-design, and precision medicine could ultimately transform herbal formulation development from an empirical, trial-and-error process into a predictive, mechanistically informed, and experimentally validated discipline.

REFERENCES

1. Khan, S. R., Al Rijjal, D., Piro, A., & Wheeler, M. B. (2021). Integration of AI and traditional medicine in drug discovery. *Drug discovery today*, *26*(4), 982-992.
2. Maciejewska-Turska, M., Georgiev, M. I., Kai, G., & Sieniawska, E. (2025). Advances in bioinformatic methods for the acceleration of the drug discovery from nature. *Phytomedicine*, *139*, 156518.
3. Othman, Z. K., Ahmed, M. M., Kasimieh, O., Musa, S. S., Branda, F., Cue, E. G., ... & Vimolmangkang, S. (2025). Artificial intelligence for natural product drug discovery and development: current landscape, applications, and future directions. *Intelligence-Based Medicine*, 100316.
4. Xie, S., Zhan, F., Zhu, J., Xu, S., & Xu, J. (2025). The latest advances with natural products in drug discovery and opportunities for the future: a 2025 update. *Expert Opinion on Drug Discovery*, *20*(7), 827-843.
5. Yuan, X., Yang, X., Pan, Q., Luo, C., Luan, X., & Zhang, H. (2025). Artificial intelligence in natural products research. *Chinese Journal of Natural Medicines*, *23*(11), 1342-1357.
6. Xu, T., Xu, Y., Zhang, J., Zhou, Y., Feng, H., Zhang, A., & Zhang, Y. (2025). Advances in

AI for predicting pharmacological properties of natural medicines. *Life sciences*, 124180.

7. Gormley, A. J. (2024). Machine learning in drug delivery. *Journal of Controlled Release*, *373*, 23-30.
8. Feng, Y., Zhu, X., & Wang, Y. (2025). Application of spectroscopic technology with machine learning in Chinese herbs from seeds to medicinal materials: The case of genus Paris. *Journal of Pharmaceutical Analysis*, *15*(2), 101103.
9. Ratnasekhar, C. H., Rai, A. K., Rakwal, P., Khan, S., Verma, A. K., Mukhopadhyay, P., ... & Trivedi, P. K. (2025). 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*, *482*, 144078.
10. Wu, P., Chen, L., Xu, Y. X., Ren, Y. Y., Dong, X. M., Fan, J. X., ... & Liu, Z. Q. (2025). Decoding herbal medicine: AI-powered omics and network pharmacology. *Phytomedicine*, *148*, 157453.
11. Sridhar, D., Manikandan, R., Dhanapal, Y., Khute, S., & Subash, P. (2025). Artificial intelligence in predicting personalized nanocarrier formulations for herbal drugs: bridging phytomedicine and precision nanotechnology. *Intelligent Pharmacy*.
12. Shen, C., Zhang, M., Lu, M., Chang, E., Gao, Z., Ban, W., ... & Jiang, C. (2025). Machine learning empowered formulation design, optimization and characterization of nanoparticulate drug delivery systems: Current applications, challenges, and future perspectives. *Acta Pharmaceutica Sinica B*.
13. Liu, X., Liu, X., Wang, J., Zang, D., Yang, Y., Chen, Q., & Guo, D. A. (2025). Machine learning and chemometric methods for high-throughput authentication of 53 Root and Rhizome Chinese Herbal using ATR-FTIR fingerprints. *Journal of Chromatography B*, *1260*, 124630.
14. Guo, Q., Li, M. X., Fu, R., Wan, X., Dong, W. H., Mao, C. Q., ... & Li, Y. (2025). Rapid evaluation of Curcuma origin and quality based on E-eye, flash GC e-nose, and FT-NIR combined with machine learning technologies. *Food Chemistry*, *481*, 143953.
15. Talebi, M., Shahbazi, K., Dakkali, M. S., Akbari, M., Ghale, R. A., Hashemi, S., ... & Aminzadeh, S. (2025). Phytosomes: A promising nanocarrier system for enhanced bioavailability and therapeutic efficacy of herbal products. *Phytomedicine Plus*, *5*(2), 100779.
16. Shaw, I., Aryee, A. A., Ali, Y. S., Boafo, G. F., Tian, J., Mlambo, R., ... & Chen, C. (2025). Harmonizing tradition and technology: Liposomal nanocarriers unlocking the power of natural herbs in Traditional Chinese Medicine. *Chinese journal of natural medicines*, *23*(6), 700-713.
17. Wu, Y., Wang, N., Xiong, P., Wang, R., Deng, J., & Ouyang, D. (2025). Artificial intelligence for drug delivery: Yesterday, today and tomorrow. *Acta Pharmaceutica Sinica B*.
18. Zulbeari, N., Wang, F., Mustafova, S. S., Parhizkar, M., & Holm, R. (2025). Machine learning strengthened formulation design of pharmaceutical suspensions. *International Journal of Pharmaceutics*, *668*, 124967.
19. Das, P., Halder, S., Chowdhury, D., Palai, S., Goswami, S., Ghosh, A., ... & Naskar, S. (2026). Harnessing Artificial Intelligence in Unravelling Dissolution Profiling of Herbal Therapeutics: Transformative Prospects, Inherent Challenges and Methodological Boundaries. *Current Pharmacogenomics and Personalized Medicine*, *23*(1), E18756921429784..

20. Noorain, Srivastava, V., Parveen, B., & Parveen, R. (2023). Artificial intelligence in drug formulation and development: applications and future prospects. *Current drug metabolism*, *24*(9), 622-634.
21. Zhu, T., Liu, B., Chen, N., Liu, Y., Wang, Z., & Tian, X. (2025). Artificial intelligence-driven innovations in pharmaceutical development and drug delivery systems. *Current topics in medicinal chemistry*, *25*(25), 2937-2951.
22. Kaur, R., Baldi, A., & Baldi, A. (2025). Artificial Intelligence in Herbal Drug Authentication: Revolutionizing Identification, Adulterant Detection and Standardization. *Current Computer Science*, *4*(1), E29503779378012.
23. Arunraj, K. P., Haritha, K. M., Khulood, M. T., Sana, P. A., Thanha, K. P. K., & Pramod, K. (2025). The Use of Artificial Intelligence in the Formulation of Effervescent Tablets: A Review. *Mini-Reviews in Medicinal Chemistry*.
24. Arora, S., Chettri, S., Percha, V., Kumar, D., & Latwal, M. (2024). Artifical intelligence: a virtual chemist for natural product drug discovery. *Journal of Biomolecular Structure and Dynamics*, *42*(7), 3826-3835.
25. Mazumdar, H., Khondakar, K. R., Das, S., Halder, A., & Kaushik, A. (2025). Artificial intelligence for personalized nanomedicine; from material selection to patient outcomes. *Expert Opinion on Drug Delivery*, *22*(1), 85-108.
26. Favaron, A., Abdalla, Y., Basit, A. W., & Orlu, M. (2025). Artificial Intelligence's role in colon-targeted drug delivery. *Expert Opinion on Drug Delivery*, *22*(3), 315-319.
27. Filer, C. N. (2025). Artificial intelligence and natural product research. *Natural Product Research*, *39*(7), 2051-2053.
28. Akhtar, M., Nehal, N., Gull, A., Parveen, R., Khan, S., Khan, S., & Ali, J. (2025). Explicating the transformative role of artificial intelligence in designing targeted nanomedicine. *Expert Opinion on Drug Delivery*, *22*(7), 971-991.

HOW TO CITE: Alisha Jabi, Deepika Shakya, Priti Yadav, Artificial Intelligence-Assisted Herbal Formulation Development: From Phytochemical Intelligence to Predictive Nanodelivery and Precision Phytotherapy, Int. J. of Pharm. Sci., 2026, Vol 4, Issue 9, 2357-2373, https://doi.org/10.5281/zenodo.22868186