

Research Insight

Open Access

Computational Analysis of Growth Characteristics and Active Compound Accumulation in Zhejiang Medicinal Plants

Weiduo Liu^{1,2} ✉¹ Zhejiang zhongxin Medical Plant Seed Co., Ltd, Hangzhou, 311402, Zhejiang, China² Zhejiang Agronomist College, Hangzhou, 310021, Zhejiang, China✉ Corresponding author: 454006383@qq.comGenomics and Applied Biology, 2026, Vol.17, No.5 doi: [10.5376/gab.2026.17.0024](https://doi.org/10.5376/gab.2026.17.0024)

Received: 22 Aug., 2026

Accepted: 26 Sep., 2026

Published: 11 Oct., 2026

Copyright © 2026 Liu, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:

Liu W.D., 2026, Computational analysis of growth characteristics and active compound accumulation in Zhejiang medicinal plants, Genomics and Applied Biology, 17(5): 312-325 (doi: [10.5376/gab.2026.17.0024](https://doi.org/10.5376/gab.2026.17.0024))

Abstract Zhejiang Province is recognized as one of the important regions for medicinal plant diversity in China, with abundant resources characterized by unique ecological adaptability and rich accumulation of bioactive compounds. However, the growth performance and medicinal quality of these plants are strongly influenced by complex interactions among genetic background, environmental conditions, cultivation practices, and metabolic regulation processes. Recent advances in computational biology, artificial intelligence, and multi-omics technologies provide new opportunities for systematically investigating plant growth dynamics and active compound accumulation. This review summarizes the biological characteristics and developmental patterns of Zhejiang medicinal plants and discusses computational frameworks integrating phenotypic data, environmental parameters, metabolomics, transcriptomics, and machine learning approaches. Particular emphasis is placed on computational modeling strategies for predicting growth traits, identifying key regulatory factors, and elucidating metabolic pathways associated with bioactive compound biosynthesis. Furthermore, advanced technologies, including remote sensing-based digital phenotyping, network biology, and deep learning models, are evaluated for their potential applications in precision cultivation and medicinal quality optimization. A case study framework is presented to demonstrate how machine learning models can integrate environmental variables, growth indicators, and metabolite profiles to predict plant productivity and medicinal compound accumulation. Despite significant progress, challenges remain in data standardization, model interpretability, and the integration of computational predictions with experimental validation. Future development of artificial intelligence-driven platforms, digital twins, and multi-scale biological models will facilitate sustainable cultivation, resource conservation, and quality improvement of Zhejiang medicinal plants. This computational perspective provides a theoretical and technological foundation for advancing intelligent medicinal plant research and precision herbal medicine production.

Keywords Zhejiang medicinal plants; Computational biology; Machine learning prediction; Active compound accumulation; Multi-omics integration

1 Introduction

Medicinal plants form the material foundation of traditional Chinese medicine and remain increasingly important for health care, industry, and regional bioeconomies. At the same time, sustainable development of this resource base is constrained by declining wild populations, uneven cultivation practices, and persistent variation in the quality of cultivated materials. These challenges are especially relevant in provinces such as Zhejiang, where medicinal-plant richness, industrial potential, and geographic heterogeneity coexist, yet development of planting systems, markets, and enterprise distribution remains imperfectly aligned with resource endowment (Shan et al., 2021). Environmental variation further complicates production because soil, climate, and terrain strongly influence plant growth and the accumulation of active ingredients, often causing substantial differences within the same species across habitats and medicinal parts. For Zhejiang medicinal plants, this means that evaluating growth vigor alone is insufficient; robust assessment must also address phytochemical quality, environmental responsiveness, and the interaction between agronomic performance and medicinal value.

Recent empirical work has reinforced the importance of linking growth traits with bioactive compound accumulation through quantitative analysis. In *Polygonatum cyrtoneura*, a medicinal and edible species of high economic relevance in southern China, multi-site provenance testing in Zhejiang showed highly significant

differences among provenances, sites, and provenance-by-site interactions for all measured growth and medicinal traits. The same study reported broad phenotypic and genotypic variation, high repeatability for many traits, and positive associations between growth characteristics and key medicinal components, indicating strong potential for selection and breeding. Geo-climatic analyses further suggested that metabolite accumulation is not random: saponin content increased with altitude and temperature, whereas flavonoid accumulation was promoted under drier conditions (Cheng et al., 2026). Such findings underscore a central problem for medicinal plant research in Zhejiang: growth traits, environmental adaptation, and active compound accumulation are tightly coupled, but their relationships are multidimensional and difficult to resolve using conventional descriptive methods alone.

Advances in computational approaches now provide a practical framework for addressing this complexity. Reviews of medicinal plant informatics show that computational research has expanded from molecular docking and molecular dynamics simulation to artificial intelligence, including artificial neural networks, deep neural networks, and related machine-learning tools that reduce the time and cost of conventional screening. More recent work emphasizes that machine learning is particularly valuable when medicinal plant quality must be inferred from multi-source data, because single analytical techniques often fail to capture the multi-component nature of medicinal materials and their environmental responsiveness. Parallel advances in sequencing and analytical chemistry have also widened the computational toolbox: RNA-seq and related technologies can identify genes involved in biosynthetic pathways, while spectroscopy and chromatography generate the metabolite-level data needed for integrated modeling (Singh et al., 2022). At a higher systems level, emerging approaches such as AI-driven multi-omics integration and digital twin modeling seek to connect real-time or high-dimensional data with predictive models of growth dynamics and metabolite yield, although challenges remain in causal inference, interpretability, and cross-species generalization (Chen et al., 2026).

Against this background, this review aims to synthesize current knowledge on how computational analysis can be used to characterize growth characteristics and active compound accumulation in Zhejiang medicinal plants, with particular attention to the integration of ecological, phenotypic, phytochemical, and multi-omics information. The conceptual framework adopted here treats medicinal plant quality as an emergent property of interacting biological and environmental layers, rather than as a single trait measured in isolation (Chen et al., 2026). Accordingly, the review is organized around three connected objectives: first, to summarize the sources of variation in growth performance and secondary metabolite accumulation in Zhejiang-relevant medicinal plants; second, to examine computational methods capable of modeling these variations across scales, from trait statistics to network and AI-based prediction; and third, to identify how such methods can support germplasm evaluation, cultivation optimization, quality control, and sustainable resource use (Cheng et al., 2026). By combining regional cultivation evidence with advances in computational biology, this review positions Zhejiang medicinal plants as a model system for data-driven modernization of medicinal plant research and industry.

2 Biological Characteristics and Growth Dynamics of Zhejiang Medicinal Plants

2.1 Diversity, distribution, and habitat adaptation

Zhejiang medicinal plants should be understood within the broader biogeographic pattern of Chinese medicinal flora, in which richness is concentrated in central and southern China and shaped by fine-scale environmental heterogeneity. High-resolution national mapping showed that a small fraction of grid cells contains most medicinal plant diversity, while endemic medicinal plant richness is positively associated with contemporary precipitation and altitudinal range, both of which are highly relevant to Zhejiang's humid subtropical and topographically varied landscapes. This pattern suggests that Zhejiang's mountainous terrain, dissected watersheds, and strong local climatic gradients likely promote species coexistence, ecological specialization, and habitat partitioning among medicinal taxa rather than supporting a uniform regional flora.

Habitat adaptation in Zhejiang also depends on how populations respond to fragmentation, geographic isolation, and local environmental filters. A Zhejiang-focused population genomic study of *Coptis chinensis* var. *brevisepala* identified Zhejiang as the core distribution area of the species and found strong population differentiation largely driven by geographic isolation, with ultraviolet radiation and low temperature contributing to fine-scale

divergence. Complementary evidence from subtropical mountain forests in nearby Fujian showed clear elevational replacement of medicinal plant communities and stronger environmental effects on shrub and herb β -diversity than on tree-layer diversity, indicating that medicinal plant habitat adaptation in southeastern China is strongly stratified across elevation and vegetation layers.

2.2 Growth phenology and developmental characteristics

The growth dynamics of Zhejiang medicinal plants are inherently seasonal, and their developmental trajectories are best interpreted through standardized phenological frameworks. Broad phenological theory shows that plants time leafing, flowering, fruiting, and dormancy in ways that reflect adaptation to local environments, while long-term phenology linked with molecular and climate data can improve prediction of future shifts. For medicinal plant research, this matters because the timing and duration of developmental stages determine not only biomass accumulation but also the windows during which organs with pharmacological value are formed, mature, and harvested.

Standardized BBCH-based studies provide practical models for describing these developmental sequences in medicinal species. In *Astragalus membranaceus* var. *mongholicus*, phenological development was organized into distinct vegetative, reproductive, and senescence stages, and the chronology of these stages was linked to accumulated thermal time through growing degree days. Mechanistic developmental evidence from first-year *Panax ginseng* further showed that storage-root thickening follows a stage-specific program in which early growth is associated with auxin, gibberellin, and nitrate signaling, middle growth with cell division and wall biogenesis, and late growth with jasmonic-acid-associated preparation for dormancy.

2.3 Environmental regulation of growth performance

Environmental regulation of medicinal plant growth in Zhejiang is likely to be multidimensional, because temperature, light, water, soil conditions, and nutrient supply act together on both growth and phytochemical traits. Reviews across medicinal species show that environmental optimization can substantially improve plant growth and metabolite synthesis, while ecologically limiting factors such as temperature, lighting, soil water, salinity, and fertility alter both physiological performance and secondary metabolism (Pant et al., 2021). For Zhejiang production systems, this means that growth performance cannot be evaluated independently of environmental management, especially where high humidity, seasonal rainfall, and mountainous microclimates create sharp local differences in resource availability and stress exposure.

More recent cultivation studies indicate that environmental effects are not only strong but also highly context dependent. Controlled-environment systems can improve the consistency, concentration, and yield of bioactive phytochemicals by precisely regulating light, carbon dioxide, temperature, humidity, nutrients, and airflow, whereas climate-related stresses can reshape plant morphology, physiology, and secondary metabolite production in less predictable ways under field conditions. Experimental evidence also shows that microclimate manipulation can shift growth-quality tradeoffs: in medicinal crops grown under dynamic agrivoltaics, heavy shade reduced biomass in most species but increased essential oil yield and sharply lowered evapotranspiration, highlighting the need for computational models that optimize both growth performance and active compound accumulation rather than maximizing yield alone.

3 Computational Frameworks for Growth Trait Analysis in Medicinal Plants

3.1 Data acquisition and multi-source dataset integration

Computational analysis of growth traits in Zhejiang medicinal plants depends first on building datasets that combine field observations with remote sensing and chemical measurements. In medicinal plants, growth and quality vary strongly with environment and across medicinal parts, so single-source measurements often fail to capture the biological basis of active-compound accumulation. Reviews of medicinal-plant monitoring therefore emphasize integrated satellite-, UAV-, and ground-based sensing as the main route toward dynamic growth retrieval, stress monitoring, and non-destructive quality evaluation. At the operational level, multi-source acquisition increasingly joins spectral imagery with synchronous phenotypic sampling. In *Glycyrrhiza uralensis*, UAV multispectral images were collected together with plant height, tiller number, SPAD, and nitrogen content,

showing how field and aerial observations can be aligned into one growth-monitoring framework (Zhang et al., 2025). More broadly, multimodal remote sensing improves crop monitoring because different sensing modalities complement one another and provide a more comprehensive representation of plant growth than single devices alone.

For Zhejiang medicinal plants, dataset integration should also be spatially explicit and cultivation-oriented. Machine learning combined with GIS can organize soil, subregion, and medicinal-plant information to identify suitable cultivation zones and support targeted management strategies (Roopashree et al., 2024). At the same time, UAV-based high-resolution imaging enables plant-level monitoring across the growing season, making it possible to estimate maturity and detect growth problems early enough for intervention (Vigneault et al., 2023). A further development direction is extending data integration beyond morphology into biochemical and omics layers. AI-driven multi-omics workflows are designed to integrate heterogeneous biological datasets so that fragmented information can be assembled into metabolic networks relevant to compound biosynthesis (Figure 1) (Chen et al., 2026). This is especially important in medicinal plants because multi-source data fusion has been shown to improve holistic quality evaluation and prediction compared with single analytical methods.

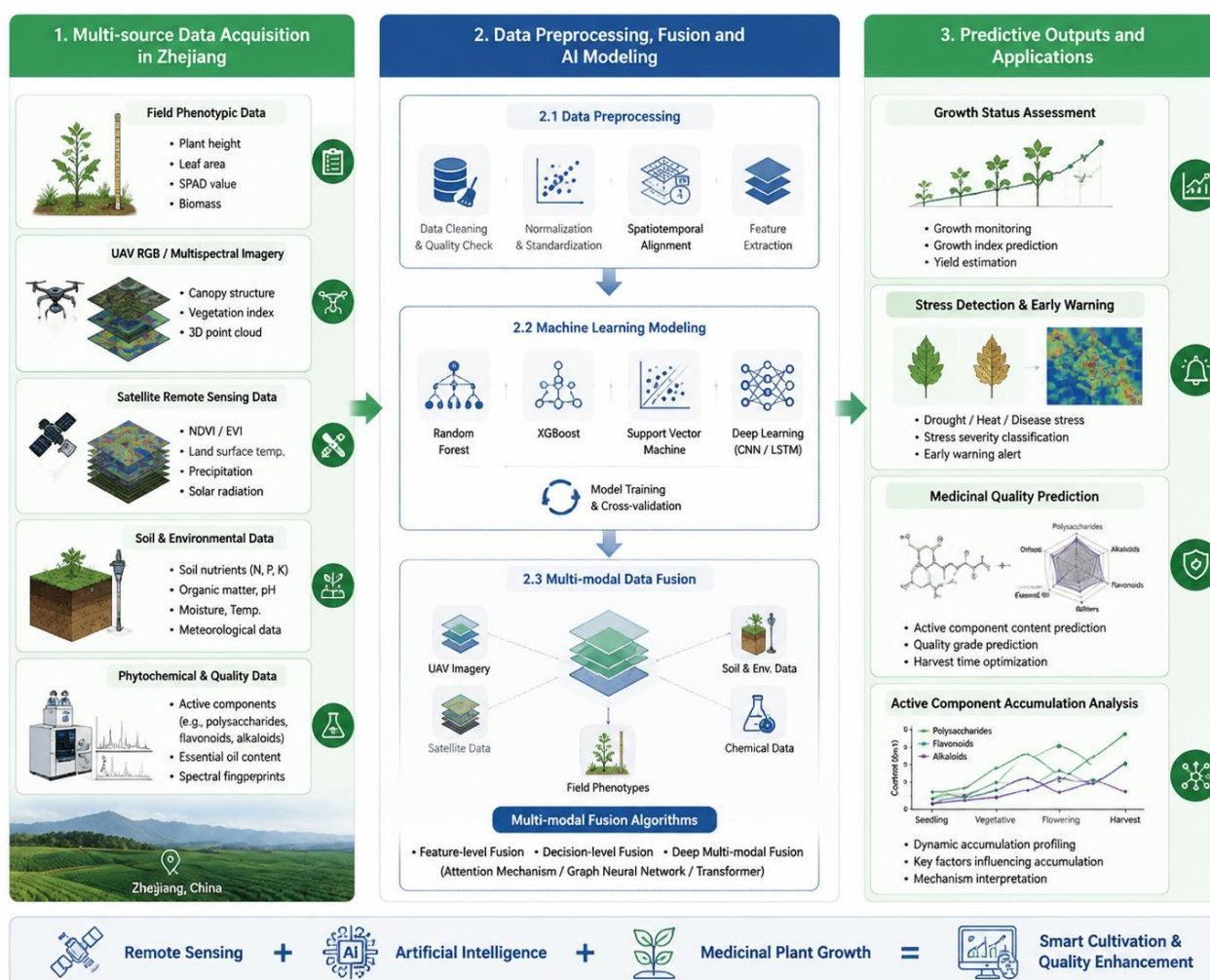


Figure 1 Multi-source data integration framework for computational analysis of growth traits and quality evaluation in Zhejiang medicinal plants

3.2 Machine learning and statistical modeling of growth characteristics

Once multi-source datasets are established, machine learning and statistical models become the core tools for extracting growth patterns and trait relationships. In medicinal plants, these methods help analyze the joint effects

of genetics, environment, and cultivation conditions, and can support decisions on harvesting time, irrigation, fertilization, and genotype selection (Roopashree et al., 2024). More generally, plant phenotyping research shows that machine learning can reduce the bottleneck created by large high-throughput datasets by linking extracted features to measurable phenotypes through classification, regression, and prediction. Model choice depends strongly on sample size, task complexity, and the need for interpretability. Traditional approaches such as PLSR, SVM, and decision-tree methods remain useful when training data are limited because they are interpretable and computationally modest, although they rely on manual feature engineering and capture nonlinear structure less effectively. By contrast, neural-network-based models and related modern phenomics frameworks have improved the precision of long-term forecasting and spatiotemporal pattern recognition in plant-growth prediction tasks (Debbagh et al., 2025).

Recent empirical studies confirm that nonlinear models can accurately estimate medicinal-plant growth indicators. In *Glycyrrhiza uralensis*, BP, SVM, and RF models were used to predict phenotypic indicators and yield, and integrated modeling with multiple indicators substantially improved performance (Zhang et al., 2025). Similar advantages appear in broader crop phenotyping, where deep learning models trained on time-series imagery have predicted future root and shoot growth with performance close to expert annotation and adaptability across plant species. Statistical and machine-learning models are also increasingly used to connect growth traits with metabolite accumulation. ANN-based and neurofuzzy approaches have been proposed as effective tools for modeling multifactorial processes that influence phenolic-compound production, while improving interpretability through rule-based simplification (García-Pérez et al., 2020). Reviews of genotype-to-phenotype prediction likewise argue that machine learning can outperform conventional statistical tools in high-dimensional settings because it can extract features from genetic, environmental, and image data more flexibly.

3.3 Artificial intelligence-based prediction and decision support systems

Artificial-intelligence-based prediction systems extend trait analysis into practical decision support by converting monitoring results into cultivation recommendations. In medicinal-plant production, this is especially valuable because AI can optimize not only biomass growth but also the accumulation of pharmacologically important secondary metabolites across variable conditions (Chen et al., 2026). Recent reviews of medicinal herb breeding therefore frame AI as a platform technology spanning multi-omics integration, trait optimization, intelligent monitoring, and genotype-environment-management interaction. At the farm-management level, decision support is moving from empirical adjustment toward real-time, demand-oriented control. UAV-guided monitoring in *Glycyrrhiza uralensis* was used to support variable irrigation and nitrogen topdressing, illustrating how integrated phenotyping can inform immediate water-nutrient management decisions (Zhang et al., 2025). Comparable work in plant-scale UAV monitoring has shown that logistic growth curves and dashboard-based reporting can project maturity before harvest and guide field interventions under heterogeneous conditions (Vigneault et al., 2023).

A second major application is predictive optimization of active-compound production. AI-based metabolite research shows that ANN models can learn nonlinear relationships among growth regulators, nutrient composition, light, elicitors, and metabolite accumulation, allowing prediction of optimal culture conditions without exhaustive experiments. This shift is important because conventional in vitro metabolite production is often irreproducible due to genetic variability, environmental fluctuation, and pathway regulation, all of which favor data-driven optimization (Srivastava and Bharadvaja, 2026). Future intelligent systems for Zhejiang medicinal plants will likely combine explainable prediction with broader biological and agronomic knowledge. Reviews of AI in medicinal-plant research emphasize not only phytochemical profiling and predictive modeling, but also the integration of AI with traditional ethnobotanical expertise to improve agricultural output and conservation. At the same time, future phenomics frameworks are expected to benefit from tighter integration of domain knowledge with data-driven methods and from more comprehensive standardized datasets, which should improve robustness in real cultivation settings (Debbagh et al., 2025).

4 Computational Analysis of Active Compound Accumulation Mechanisms

4.1 Biosynthetic pathways and regulatory networks of bioactive compounds

Computational analysis of active compound accumulation in Zhejiang medicinal plants begins with reconstructing biosynthetic pathways and the regulatory networks that control them. Recent reviews show that omics platforms now enable comprehensive mapping of biosynthetic pathways, regulatory circuits, and spatial chemical distributions, which has made pathway discovery in medicinal plants far more systematic than earlier gene-by-gene approaches (Latif and Nawaz, 2025). This is important because medicinal metabolites are produced through complex pathways whose full elucidation is necessary for predicting metabolic switches and building sustainable production strategies (Chen et al., 2026). A central advance has been the move from simple correlation to network-based inference of regulatory control. Multi-omics studies can link transcripts, metabolites, proteins, and chromatin features into layered networks that identify co-regulated genes and likely regulatory nodes, while newer AI-enabled approaches aim to infer missing links and pathway bottlenecks more directly. Evidence from medicinal plants supports this framework: transcriptomic analysis in *Cibotium barometz* identified tissue-specific key enzymes and transcription-factor correlations for lignin and flavonoid pathways, while integrated profiling in *Trichosanthes kirilowii* highlighted Tk_ERF4 and related regulators as coordinators of tissue- and stage-specific bioactive compound biosynthesis (Zhang et al., 2025; Gao et al., 2026).

4.2 Metabolomics and multi-omics integration for compound prediction

For compound prediction, metabolomics serves as the most direct layer because metabolites are the executors of gene function and the final products of pathway activity. Metabolomics combined with other omics can reveal metabolite-gene associations at the whole-genome level, identify enzymes that control production or transformation, and provide a systems-level basis for predicting which compounds accumulate in specific tissues or developmental stages. In medicinal plants, this integrated strategy is increasingly viewed as the most effective route to uncover complex specialized-metabolite pathways that cannot be resolved by single-omics data alone.

Prediction becomes more powerful when integration is spatiotemporal rather than static. Transcriptomics and metabolomics provide complementary information by associating temporal and spatial gene-expression patterns with metabolite abundance, and time-series regulatory analysis is especially useful for resolving dynamic pathway components (Singh et al., 2022). Recent medicinal-plant work confirms this value: integrated multiomics in *Trichosanthes kirilowii* resolved tissue-specific accumulation of cucurbitacin B, terpenoids, and flavonoids and identified the color-changing fruit stage as a critical window for secondary metabolism, while spatial multi-omics in other systems is expected to clarify synthesis, transport, and accumulation at cell-level resolution (Wu et al., 2025; Gao et al., 2026).

4.3 Environmental and agronomic regulation of active compound accumulation

Active compound accumulation is also shaped by environmental and agronomic regulation, so computational models must account for external drivers rather than treating metabolism as genetically fixed. Broad reviews show that secondary-metabolite accumulation responds strongly to light, temperature, soil water, fertility, salinity, and other environmental factors, and even a change in one factor can alter metabolite content under otherwise similar conditions (Pant et al., 2021). This sensitivity explains why prediction models for Zhejiang medicinal plants should incorporate field microclimate, soil status, and developmental context when estimating active compound accumulation.

Mechanistically, environmental regulation acts through stress and signaling networks that alter biosynthetic gene expression and metabolite flux. Stress conditions elevate reactive oxygen species, activate MAPK cascades, and trigger hormonal pathways involving salicylic acid, jasmonic acid, ethylene, and abscisic acid, which then regulate transcription factors and biosynthetic modules linked to specialized-metabolite accumulation. Light is one of the clearest examples: different light qualities, intensities, and photoperiods modulate pathway activity through photoreceptor-mediated signaling and circadian regulation, making light management a plausible lever for directed enhancement of medicinal compounds under controlled or precision cultivation systems (Wu et al., 2025).

5 Advanced Computational Technologies for Zhejiang Medicinal Plant Research

5.1 Remote sensing and digital phenotyping technologies

Remote sensing and digital phenotyping are becoming core technologies for medicinal plant research because they replace labor-intensive, subjective field inspection with scalable and dynamic observation. Reviews of medicinal plant cultivation show that remote sensing addresses major limits of conventional methods, while broader assessments indicate that integrated satellite-, UAV-, and ground-based sensing can already support distribution mapping, growth retrieval, stress monitoring, and non-destructive quality evaluation across many medicinal species. For Zhejiang medicinal plants, this is especially valuable because cultivation systems require repeated measurements of canopy condition, habitat variation, and developmental stage over large and heterogeneous production areas. At the field scale, UAV-based phenotyping has shown that medicinal plant growth can be monitored with high precision when spectral indices are combined with direct phenotypic measurements. In *Glycyrrhiza uralensis*, high-resolution full-phenology UAV data supported BP, SVM, and RF models for yield forecasting, and combining multiple indicators improved accuracy beyond single-parameter prediction (Zhang et al., 2025). Similar progress appears in wild-resource surveys, where UAV imagery plus Mask R-CNN enabled automated identification, counting, yield prediction, and spatial mapping of individual *Lamioophlomis rotata* plants in complex high-altitude environments (Ding et al., 2023).

Digital phenotyping is also advancing from simple imaging toward environmentally informed monitoring. Remote sensing can characterize growth-regulating variables such as soil moisture and soil-available silicon, and long-term time-series observation can track medicinal-plant phenology together with climate responses. This matters because medicinal plant phenotypes emerge from genotype-environment interaction, so Zhejiang-focused digital phenotyping platforms should link plant traits with water, nutrient, soil, and weather data rather than treating canopy images as isolated inputs. A further technological shift is the move toward chemical-sensitive phenotyping rather than morphology alone. Hyperspectral imaging captures continuous spectral information related to chemical composition and has supported both traditional machine learning and deep learning for component inversion, including simultaneous prediction of multiple ginsenosides through attention-based temporal convolutional models. Because medicinal plant quality depends on the joint distribution of multiple active constituents, these methods are likely to be especially important for Zhejiang geo-authentic herbs where growth monitoring and quality monitoring must be integrated within the same sensing framework.

5.2 Network biology and systems-level modeling

Network biology provides the systems framework needed to connect growth characteristics with active compound accumulation. Plant metabolic network modeling has emerged as a major tool for integrating and predicting the spatial and temporal distribution of metabolic flows, and recent reviews emphasize that mathematical models are essential for investigating growth, development, metabolic regulation, and secondary metabolism across plant stages (Rao and Liu, 2025). For Zhejiang medicinal plants, this means that systems-level models can serve as bridges between observable traits in the field and less visible pathway-level changes that determine compound accumulation. These models are particularly valuable because medicinal plant metabolism is not a single-pathway process but a coordinated network with competing fluxes, feedback regulation, and environmental sensitivity. Network modeling is now used to quantify pathway fluxes and guide strategies for directing metabolism in plant natural product systems, while broader systems-biology perspectives argue that advances in bioinformatics are shifting medicinal plant research away from reductionist explanation toward network-based interpretation (Noor et al., 2022; Rao and Liu, 2025). This transition is important for Zhejiang studies seeking to explain why the same species shows different growth quality or active-compound profiles under different ecological and agronomic settings.

Another key application is linking chemical constituents to targets, pathways, and biological functions through network pharmacology. This approach typically starts by identifying active compounds and predicting associated genes or targets from chemical representations, providing a structured route from phytochemical data to multi-target functional hypotheses (Noor et al., 2022). In traditional Chinese medicine research more broadly, network pharmacology is used to build multilayer relationships among compounds, targets, and pathways, and it

supports the identification of efficacy-related marker components that can improve quality evaluation of complex herbal materials. Systems-level modeling is also becoming more predictive as it incorporates multi-omics and machine learning. Reviews of medicinal plant metabolic networks note that AI methods can extract key biological features from high-dimensional datasets and support the construction of digital-twin-like surrogate systems for optimizing elicitation, cultivation, and metabolic regulation (Chen et al., 2026). For Zhejiang medicinal plants, this suggests a future workflow in which transcriptomic, metabolomic, and phenotypic data are embedded in network models that can forecast how cultivation decisions alter both plant growth and secondary metabolite output (Chen et al., 2026).

5.3 Deep learning and predictive analytics in medicinal plant science

Deep learning and predictive analytics are extending medicinal plant research from description to high-dimensional inference. Recent reviews show that deep learning is particularly effective for analyzing complex data in bioactivity prediction and drug discovery, while medicinal-plant-specific assessments emphasize that deep models perform well on nonlinear multi-omics datasets and can reveal regulatory modules associated with secondary metabolite synthesis (Prajapati et al., 2025). This makes deep learning especially relevant for Zhejiang medicinal plants, where trait development and active-compound accumulation are shaped by interacting genetic, chemical, and environmental factors. One practical advantage of predictive analytics is automated feature discovery from heterogeneous data. Machine learning has been used to identify crucial metabolite features affecting phenolic synthesis, and in medicinal plant hyperspectral applications it can select characteristic wavelengths for accurate prediction of compounds such as tanshinones and for origin classification (Chen et al., 2026). Beyond spectroscopy, automated machine learning has also been used to predict precursors of specialized metabolites, with regularized linear classifiers outperforming prior methods while remaining interpretable, which is useful for accelerating pathway discovery in plant secondary metabolism.

Deep learning is also reshaping phenotyping and resource assessment through image-based prediction. In medicinal plant surveys, UAV imagery combined with deep learning has enabled high-accuracy instance segmentation of individual plants and improved yield estimation in complex field environments (Ding et al., 2023). In cultivation-oriented work, integrated UAV phenotyping has further supported real-time water-nutrient decision making, showing that predictive models can move beyond passive monitoring toward operational management support (Zhang et al., 2025). Despite these advances, predictive analytics in medicinal plant science still faces limits in interpretability, transferability, and data standardization. Deep learning models often act as black boxes and may not provide biologically verifiable mechanisms, while broader reviews of natural-compound modeling identify the need for benchmark datasets, interpretability tools, and stronger experimental validation (Prajapati et al., 2025). For Zhejiang medicinal plant research, the most promising direction is therefore not deep learning alone, but explainable predictive systems that combine remote sensing, multi-omics, and network knowledge into robust decision-support tools for cultivation and quality control (Prajapati et al., 2025; Chen et al., 2026).

6 Case Study: Computational Analysis of Growth Characteristics and Active Compound Accumulation in Zhejiang Medicinal Plants

6.1 Case selection and experimental dataset construction

A Zhejiang case study is best designed around representative medicinal species with clear economic value, measurable growth traits, and chemically defined active constituents. Case selection should also reflect the fact that medicinal plant quality depends on multi-component composition and varies across medicinal parts and environmental conditions, so dataset construction must capture both growth and compound heterogeneity rather than treating plants as uniform samples. This justifies choosing species for which phenotypic, ecological, and phytochemical data can be collected simultaneously, because multi-source data provide a more complete basis for quality evaluation than single analytical streams.

The experimental dataset should combine field observations, imaging, and chemical profiling in a unified sampling framework. A practical template comes from medicinal plant phenotyping studies that pair UAV or

hyperspectral measurements with synchronous ground traits such as biomass, SPAD, or morphological indicators, and that improve performance further by fusing spectral, texture, and phenotypic variables (Wu et al., 2025). Spatial layers should also be included, because machine learning linked with GIS and soil information can map suitable cultivation zones and rapidly organize medicinal plant growth conditions at regional scale, which is directly relevant for Zhejiang's heterogeneous production environments (Roopashree et al., 2024).

6.2 Machine learning-based prediction of growth performance

For growth-performance prediction, the case study should model biomass, vigor, and relative growth rate using supervised learning algorithms trained on integrated temporal and environmental features. Evidence across plant systems shows that machine learning handles complex functional associations in biomass prediction well, especially when datasets include weather, soil, and agronomic variables rather than image features alone. Ensemble methods are often strong performers in such settings, and agriculturally important predictors repeatedly include temperature, precipitation, slope, elevation, and other environmental drivers that can be measured or inferred for medicinal plant fields (Figure 2) (Cacho et al., 2023).

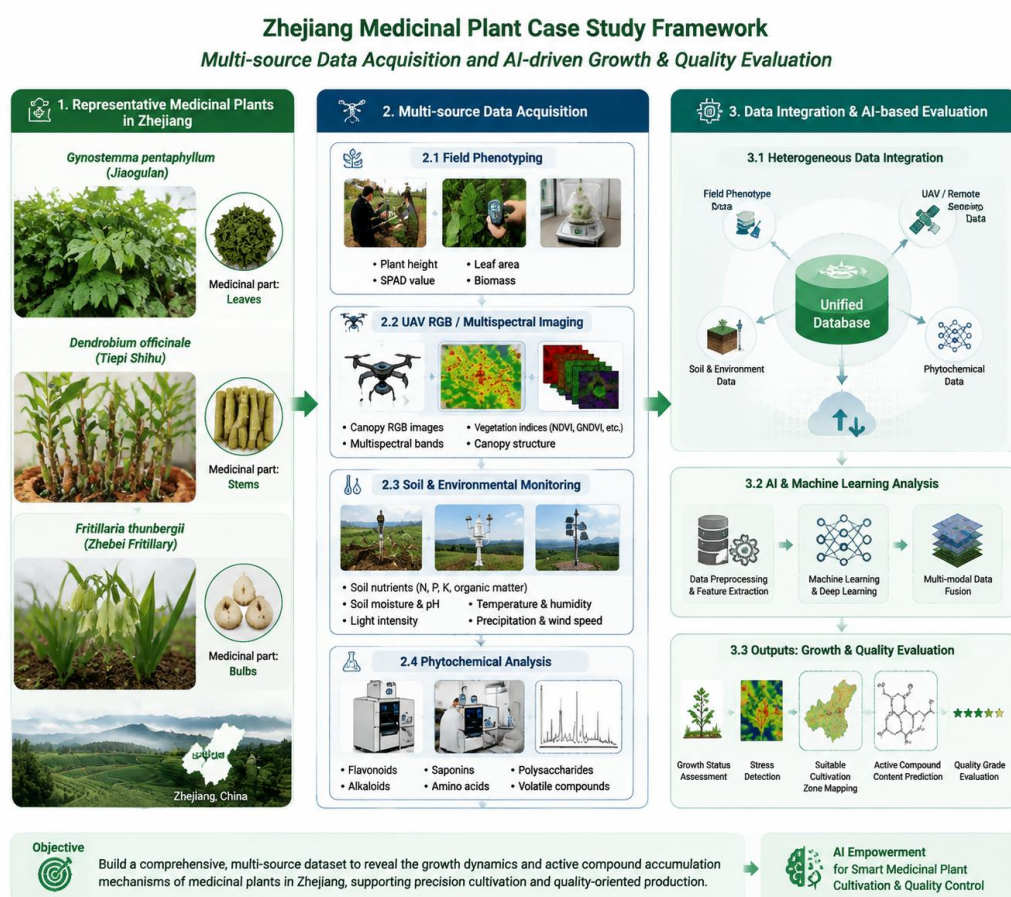


Figure 2 Representative medicinal plant selection and multi-source data acquisition framework for Zhejiang case studies

At the same time, medicinal-plant-specific evidence supports non-destructive and early prediction workflows. Hyperspectral and deep learning approaches have estimated SPAD and biomass in *Lamiophlomis rotata* with good accuracy, while feature fusion improved prediction over single feature types, indicating that Zhejiang medicinal plants would benefit from models that merge spectral and phenotypic descriptors (Wu et al., 2025). Time-resolved prediction is also valuable because early models can identify low-yield batches well before harvest and enable corrective intervention, as shown by cultivation studies where final biomass was predicted by day 8 and process adjustment increased yield by 54.1%.

6.3 Prediction of active compound accumulation and quality optimization

Prediction of active compound accumulation should be based on integrated chemical, biological, and

environmental descriptors rather than concentration values alone. Multi-omics and AI studies show that large heterogeneous datasets can be fused to identify key biosynthetic components, model dynamic metabolic behavior, and support optimization from cultivation conditions to extraction parameters (Chen et al., 2026). This is especially appropriate for Zhejiang medicinal plants because modern omics platforms can map biosynthetic pathways, regulatory networks, and spatial chemical distributions, making compound prediction more mechanistic and less empirical (Latif and Nawaz, 2025).

For quality optimization, the most effective framework is likely to combine spectrum-effect modeling with predictive control of compound content. AI-driven quality-control research in traditional Chinese medicine has shown that fused chromatographic fingerprints and learning models such as PLSR, BP-ANN, CNN, and sequence-attention networks can identify active compounds and then predict their contents non-destructively from near-infrared data (Gao et al., 2026). Complementary optimization studies also show that nonlinear and piecewise linear models can predict total polyphenols and antioxidant activity with high determination coefficients, indicating that Zhejiang case studies can use comparable regression frameworks to link cultivation and processing conditions with stable medicinal quality.

7 Challenges and Future Perspectives

7.1 Current limitations of computational modeling in medicinal plant studies

A central limitation of computational modeling in medicinal plant studies is the heterogeneity of input data. Multi-omics datasets are generated across different platforms and often carry batch effects, scale differences, and technical noise, which complicates integration and weakens downstream model robustness (Chen et al., 2026). This problem is amplified by the broader underuse and fragmented interpretation of omics datasets in plant pathway research, where large volumes of potentially informative data remain insufficiently exploited for biosynthetic inference (Reinhardt et al., 2025). A second limitation is the shortage of biological and computational foundations needed for reliable prediction. Many medicinal plants still lack high-quality reference genomes, and limited sample sizes further reduce model generalizability across species, environments, and developmental stages (Chen et al., 2026). These data constraints align with a broader challenge in AI-based natural product modeling, where prediction quality depends strongly on whether test data fall within the applicability domain of the training set and on whether high-quality integrated datasets are available at all (Xue et al., 2022).

Capturing the dynamic biology of medicinal plants remains another major obstacle. Computational models often struggle to represent nonlinear temporal relationships, especially when transcriptional responses precede measurable metabolite accumulation or when long growth cycles prevent dense time-series sampling across full developmental stages (Chen et al., 2026). More generally, plant biology still underuses mechanistic modeling even though such models are needed to relate hidden mechanisms to measurable traits and to identify emergent relationships that pattern-finding methods alone can miss. Practical adoption is also limited by human and institutional barriers. Many plant biologists still face difficulty selecting appropriate modeling strategies, especially mechanistic approaches, and often lack access to collaborators with quantitative expertise (Dale et al., 2021). In medicinal plant systems specifically, AI performance still varies with species diversity, plant condition, and the scarcity of large, diverse, publicly accessible datasets, which constrains transferability beyond narrowly defined training scenarios.

7.2 Integration of Artificial Intelligence with Experimental Biology

Future progress depends on tighter AI-experiment coupling rather than treating computation and biology as separate workflows. Current reviews emphasize a persistent prediction-validation gap and argue that medicinal plant research needs iterative computation-experiment cycles in which model outputs are repeatedly tested and refined through targeted experiments. This direction is consistent with integrated omics frameworks that combine molecular networking, reaction-pair analysis, and gene-expression patterns to improve pathway discovery while also streamlining downstream validation (Reinhardt et al., 2025). The biological value of this integration is that AI can prioritize the most informative elements for experimental follow-up. Large-scale multi-omics data now provide opportunities to discover enzymes, optimize pathway components, and improve the low abundance or

unstable quality of medicinally important compounds, but these opportunities create a parallel challenge of extracting meaningful patterns from massive datasets (Zhang et al., 2025). AI is well suited to this role because it can identify key biological components in biosynthetic systems and model metabolic behavior over time, thereby guiding experiments toward the most plausible regulatory bottlenecks (Chen et al., 2026).

The next stage is to embed AI into broader biotechnology pipelines rather than restricting it to post hoc analysis. In medicinal herb breeding, AI is already being positioned across multi-omics integration, synthetic biology, precision gene editing, trait optimization, and intelligent monitoring systems, suggesting a unified research architecture rather than isolated analytical tools. In parallel, AI and multi-omics are accelerating plant phenotype analysis, precision breeding, metabolite discovery, and quality control, which makes them increasingly relevant for medicinal plant experiments that seek simultaneous gains in growth performance and active compound output. Even so, integration with experimental biology will require stronger standards for interpretability and interdisciplinary design. Explainable AI is increasingly treated as necessary in natural product research, and meaningful feature design still depends on close collaboration among computational scientists, biologists, and medicinal chemistry specialists (Xue et al., 2022). Earlier computational reviews already suggested that AI shortens the time required for classical experimental strategies, but current evidence indicates that its real value lies in complementing rather than replacing wet-lab validation.

7.3 Future directions toward precision medicinal plant agriculture

The clearest future direction is toward precision medicinal plant agriculture, where sensing, automation, and data science are integrated to optimize both yield and phytochemical quality. Precision agriculture is defined as the coordinated use of sensors, machinery, information systems, and informed management to improve productivity under sustainable conditions, and medicinal plant systems are increasingly being drawn into this broader transition. For medicinal crops specifically, smart farming technologies already appear to improve yield, quality, and sustainability while supporting more consistent production under strict quality requirements (Khan et al., 2026). A major opportunity lies in connecting environmental control directly to biological mechanisms of compound formation. Precision irrigation, AI-driven nutrient management, and controlled-environment cultivation influence photosynthesis, nutrient uptake, stress response, and secondary metabolite biosynthesis, making them relevant not only for growth management but also for active compound optimization (Khan et al., 2026). More broadly, digital twin concepts suggest that real-time data, computational models, and simulation could eventually support continuous prediction of both plant growth and metabolite yield in medicinal plant production systems.

Another priority is the creation of multiscale breeding and management platforms. Conventional breeding struggles with polygenic regulation and the simultaneous optimization of multiple pharmacologically relevant traits across variable environments, whereas AI-based frameworks are beginning to address these combined constraints. A promising response is the proposed genotype-environment-management interactive platform, which explicitly targets data integration, model generalization, and environmental adaptation as linked problems rather than separate technical issues. For Zhejiang medicinal plants, future development will likely depend on building transparent, multiscale systems that connect cultivation, omics, and quality standardization. Recent reviews argue that stronger computational-experimental synergy and more transparent models are essential if AI is to mature into a reproducible discipline in medicinal plant science (Chen et al., 2026). At the same time, high-throughput omics integration is expected to strengthen understanding of environmental adaptation and chemical diversity, which is critical for region-specific precision cultivation and the discovery of improved medicinal plant resources (Zhang et al., 2025).

8 Conclusions

Computational approaches have made medicinal plant research more predictive by linking large, heterogeneous datasets to biologically meaningful outputs such as biosynthetic pathway reconstruction, regulator identification, and metabolic optimization). They have also broadened the analytical scope of the field beyond static screening, as molecular docking, dynamic simulation, and artificial intelligence now jointly support faster interpretation of phytochemical function and bioactivity than traditional experimental workflows alone. A second major

contribution is the integration of multi-source evidence across chemistry, plant organs, and environmental context, which improves medicinal plant quality evaluation and better reflects the ecological regulation of active compounds. At the molecular level, newer sequencing and transcriptome-based approaches have further enabled identification of genes, miRNAs, and biosynthetic pathways underlying medicinal compound formation, giving computational medicinal botany a stronger mechanistic basis. Computational methods have also contributed directly to phytochemical discovery and optimization. Machine learning combined with *in vitro* culture has been used to identify the major factors controlling phenolic compound biosynthesis and extraction, showing that computational models can guide bioactive compound production rather than only describe it. In parallel, AI-assisted metabolomics has expanded quality assessment, metabolite variation analysis, and early detection tasks, indicating that computational tools are becoming central to the full analytical pipeline of phytochemical research.

These contributions are not limited to medicinal plant cultivation but extend to drug discovery translation. *In silico* network analysis, screening, and pharmacokinetic prediction can prioritize active phytochemicals and clarify likely mechanisms before wet-lab validation, which reduces experimental burden and accelerates downstream pharmacological studies. More broadly, computational approaches are now treated as an initial key step in natural-product-based drug discovery, especially when integrated with experimental validation and modern AI methods. For sustainable cultivation, computational modeling enables site-specific decision making by connecting medicinal plant performance to soil, climate, and ecological conditions. Machine learning with GIS and soil analysis can rapidly map suitable planting areas and optimize cultivation practices, which is particularly useful for vulnerable or region-specific medicinal herbs. Spatial prediction studies further show that environmental variables can be ranked by their influence on secondary metabolite yield, allowing cultivation zones to be selected not only for survival or biomass, but also for high-value compound accumulation (Dastres et al., 2025). This has direct implications for resource conservation and regional quality management. In geo-authentic medicinal materials, ecological conditions strongly affect active ingredient accumulation, and machine learning models can forecast the geographic distribution of high-quality compounds more systematically than scattered field sampling alone. Cultivation verification studies confirm that combining field sampling, predictive simulation, and regional validation can identify optimal production areas while supporting sustainable utilization and industrial development.

Computational methods also strengthen quality control by shifting evaluation from single indicators toward integrated chemical and biological signatures. AI can correlate chemical fingerprints with documented efficacy, helping build a data-driven framework for standardization, clarification of material basis, and analysis of synergistic effects in complex medicinal systems. This is important because medicinal plant safety and efficacy depend on multi-component quality, while current quality evaluation still often overlooks variation across different medicinal parts. Sustainability also depends on the ability to adapt cultivation systems to climate change and anthropogenic disturbance. Ensemble modeling has shown that future suitable habitats for medicinal species can shift substantially under different emission scenarios, and species-specific barcoding can complement these predictions by improving germplasm identification for cultivation planning. At the production-system level, smart hydroponic and automated monitoring approaches suggest that data-driven cultivation can improve performance while enhancing sustainability, offering useful models for controlled medicinal plant production. The next stage of computational medicinal botany will likely be defined by multiscale integration. Systems biology frameworks increasingly combine co-expression analysis, regulatory network inference, graph-based learning, and explainable AI to generate interpretable predictions for complex traits and environmental responses. This trajectory is especially important for medicinal plants because future progress depends on bridging molecular networks with field-scale phenotypes across species and environments rather than analyzing each data layer in isolation.

A parallel frontier is the development of digital twin frameworks for medicinal plant agriculture. Digital twins can create real-time virtual replicas of crops or farms to support prediction and decision-making, with clear potential for optimizing irrigation, fertilization, pest management, and individualized cultivation strategies. Early agricultural digital twin work already shows that interactive, data-linked platforms can integrate multi-scale

information and move toward closed-loop management, although automated recommendation remains an important next step. Future systems will also depend on continuous monitoring and adaptive control. Reinforcement-learning-based digital twins are emerging as promising tools for optimization, automation, and resource management in agriculture, particularly where virtual environment representations can support policy learning for dynamic decision tasks. More broadly, digital farming and twin-based systems are expected to enhance productivity and efficiency through real-time synchronization between physical farms and virtual models, creating a practical foundation for precision medicinal plant agriculture. The most credible long-term outlook is therefore human-centered and trustworthy AI rather than fully autonomous black-box prediction. Agricultural AI must integrate multimodal information, remain robust to small disturbances, and explain its outputs to domain experts if it is to be trusted in biologically and economically sensitive systems. For Zhejiang medicinal plants, next-generation computational medicinal botany will be strongest when explainable AI, multi-omics, environmental sensing, and experimental validation are combined into transparent decision-support systems for sustainable cultivation and reliable quality control.

Acknowledgments

I extend my sincere gratitude to the anonymous reviewers for their valuable and insightful comments, which have greatly strengthened this paper.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Cacho J., Feinstein J., Zumpf C., Hamada Y., Lee D.J., Namoi N., Lee D., Boersma N., Heaton E., Quinn J., and Negri C., 2023, Predicting biomass yields of advanced switchgrass cultivars for bioenergy and ecosystem services using machine learning, *Energies*, 16(10): 4168.
<https://doi.org/10.3390/en16104168>
- Chen J., Cai J., Duong H.T.Q., Bunsupa S., Han R., and Tong X., 2026, AI-driven integration and optimization of medicinal plant multi-omics metabolic networks, *Frontiers in Plant Science*, 17: 1756809.
<https://doi.org/10.3389/fpls.2026.1756809>
- Cheng S., Hu Y., Cheng Y., Qian Z., Xu X., Lei X., and Shi X., 2026, Variation analysis of growth traits and medicinal components in different provenances of *Polygonatum cyrtoneura* based on heterogeneous garden experiment, *PLOS One*, 21(4): e0346920.
<https://doi.org/10.1371/journal.pone.0346920>
- Dale R., Oswald S., Jaliha A., Laporte M., Fletcher D., Hubbard A.H., Shiu S.H., Nelson A.D., and Bucksch A., 2021, Overcoming the challenges to enhancing experimental plant biology with computational modeling, *Frontiers in Plant Science*, 12: 687652.
<https://doi.org/10.3389/fpls.2021.687652>
- Debbagh M., Sun S., and Lefsrud M., 2025, Predictive modeling, pattern recognition, and spatiotemporal representations of plant growth in simulated and controlled environments: a comprehensive review, *Plant Phenomics*, 2025: 100089.
<https://doi.org/10.1016/j.plaphe.2025.100089>
- Ding R., Luo J., Wang C., Yu L., Yang J., Wang M., Zhong S.-H., and Gu R., 2023, Identifying and mapping individual medicinal plant *Lamiophlomis rotata* at high elevations by using unmanned aerial vehicles and deep learning, *Plant Methods*, 19(1): 38.
<https://doi.org/10.1186/s13007-023-01015-z>
- Gao H., Li X., Wang C., Li Y., Liu T., Chang N., Xu Y., Wang Y., Ren Y., Zhou G., Gao W., Zeng Y., Zhao H., and Li H., 2026, Integrated multiomics profiling elucidates the spatiotemporal metabolic dynamics and regulatory networks of the bioactive components of *Trichosanthes kirilowii*, *Frontiers in Plant Science*, 17: 1735703.
<https://doi.org/10.3389/fpls.2026.1735703>
- García-Pérez P., Lozano-Milo E., Landín M., and Gallego P.P., 2020, Combining medicinal plant in vitro culture with machine learning technologies for maximizing the production of phenolic compounds, *Antioxidants*, 9(3): 210.
<https://doi.org/10.3390/antiox9030210>
- Khan S., Pathania N., Kumar P., Kumar R., Kumar J., Kumar N., and Sharma A., 2026, Smart farming approaches in medicinal plant cultivation: a review of techniques, benefits, and sustainability, *Planta*, 263(4): 92.
<https://doi.org/10.1007/s00425-026-04960-w>
- Latif R., and Nawaz T., 2026, Medicinal plants and human health: a comprehensive review of bioactive compounds, therapeutic effects, and applications, *Phytochemistry Reviews*, 25(3): 2299-2342.
<https://doi.org/10.1007/s11101-025-10194-7>

- Noor F., Qamar M.T.U., Ashfaq U., Albutti A., Alwashmi A.S., and Aljasir M., 2022, Network pharmacology approach for medicinal plants: review and assessment, *Pharmaceuticals*, 15(5): 572.
<https://doi.org/10.3390/ph15050572>
- Pant P., Pandey S., and Dall'Acqua S., 2021, The influence of environmental conditions on secondary metabolites in medicinal plants: a literature review, *Chemistry and Biodiversity*, 18(11): e2100345.
<https://doi.org/10.1002/cbdv.202100345>
- Prajapati P., Shrivastav P., Prajapati J., and Prajapati B.G., 2025, Deep learning approaches for predicting bioactivity of natural compounds, *The Natural Products Journal*, 16(3): e22103155332267.
<https://doi.org/10.2174/0122103155332267241122143118>
- Rao X., and Liu W., 2025, A guide to metabolic network modeling for plant biology, *Plants*, 14(3): 484.
<https://doi.org/10.3390/plants14030484>
- Reinhardt J.K., Craft D., and Weng J.-K., 2025, Toward an integrated omics approach for plant biosynthetic pathway discovery in the age of AI, *Trends in Biochemical Sciences*, 50(4): 311-321.
<https://doi.org/10.1016/j.tibs.2025.01.010>
- Roopashree S., Anitha J., Challa S., Mahesh T., Venkatesan V.K., and Guluwadi S., 2024, Mapping of soil suitability for medicinal plants using machine learning methods, *Scientific Reports*, 14(1): 3741.
<https://doi.org/10.1038/s41598-024-54465-3>
- Shan Z., Ye J., Hao D., Xiao P., Chen Z., and Lu A., 2022, Distribution patterns and industry planning of commonly used traditional Chinese medicinal plants in China, *Plant Diversity*, 44(3): 255-261.
<https://doi.org/10.1016/j.pld.2021.11.003>
- Singh K.S., van der Hooft J.J.J., van Wees S.M., and Medema M.H., 2022, Integrative omics approaches for biosynthetic pathway discovery in plants, *Natural Product Reports*, 39(9): 1876-1896.
<https://doi.org/10.1039/D2NP00032F>
- Srivastava P., and Bharadvaja N., 2026, Role of artificial intelligence in enhancing medicinal plant metabolites, 2026 2nd international conference on cognitive computing in engineering, communications, Sciences and Biomedical Health Informatics (IC3ECSBHI), 2026: 48-53.
<https://doi.org/10.1109/IC3ECSBHI67834.2026.11469148>
- Vigneault P., Lafond-Lapalme J., Deshaies A., Khun K., de La Sablonnière S., Filion M., Longchamps L., and Mimee B., 2024, An integrated data-driven approach to monitor and estimate plant-scale growth using UAV, *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 11: 100052.
<https://doi.org/10.1016/j.ophoto.2023.100052>
- Wu X., Zhong L., Ding R., Wang C., Chen H., Zhong S., and Gu R., 2025, Non-destructive estimation of SPAD and biomass in *Lamiophlomis rotata* using hyperspectral imaging and deep learning with DRSA-CARS feature selection, *Frontiers in Plant Science*, 16: 1640779.
<https://doi.org/10.3389/fpls.2025.1640779>
- Xue H.T., Stanley-Baker M., Kong A.W.K., Li H.L., and Goh W.W.B., 2022, Data considerations for predictive modeling applied to the discovery of bioactive natural products, *Drug Discovery Today*, 27(8): 2235-2243.
<https://doi.org/10.1016/j.drudis.2022.05.009>
- Zhang A., Guan H., Dong Z., Jia X., Xue Y., Han F., Meng L., Yu X., Wang X., and Cao Y., 2025, Integrated diagnostics and time series sensitivity assessment for growth monitoring of a medicinal plant (*Glycyrrhiza uralensis* Fisch.) based on unmanned aerial vehicle multispectral sensors, *Frontiers in Plant Science*, 16: 1612898.
<https://doi.org/10.3389/fpls.2025.1612898>



Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.