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