MANTECH Publications
International Journal of Pharmacognosy and Phytochemical Sciences
Volume 4, Issue 1, January-April 2026

AI-Based Prediction of Phytochemical Bioactivity and Toxicity

Dr. Farhad Rahman[1], Shahnaj Anjum[2]
Associate Professor[1], PG Scholar[2]
Department of Pharmacy
University of Rajshahi
Corresponding Author Email: *farhad.rahman982@gmail.com[1]*
DOI: *https://doi.org/10.5281/zenodo.19707551*

ABSTRACT

Phytochemicals represent a vast and chemically diverse source of therapeutic agents. However, experimental screening of their bioactivity and safety is time-consuming, costly and often limited by availability of pure compounds. Artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools to predict biological activity and toxicity of plant-derived molecules from chemical structure data. AI-based approaches such as quantitative structure–activity relationship (QSAR) modeling, deep learning, molecular docking integration, and network pharmacology enable rapid identification of bioactive phytochemicals and early toxicity assessment. These methods accelerate drug discovery from natural products while reducing experimental burden and ethical concerns of animal testing. This review discusses current AI techniques used for phytochemical bioactivity and toxicity prediction, available phytochemical databases, modeling strategies, and validation methods. Applications in anticancer, antimicrobial and anti-inflammatory phytochemical discovery are highlighted. Challenges such as data quality, chemical diversity, interpretability, and domain applicability are also addressed. Future perspectives include explainable AI, multi-omics integration and hybrid experimental–computational pipelines for safe and effective phytochemical-based therapeutics.

KEYWORDS: *phytochemicals, artificial intelligence, bioactivity prediction, toxicity prediction, QSAR, machine learning, natural products, drug discovery*

INTRODUCTION

Plants synthesize a wide range of secondary metabolites including alkaloids, flavonoids, terpenoids, phenolics and glycosides. These phytochemicals play ecological roles in plant defense and signaling, but also exhibit diverse pharmacological activities in humans. Many modern drugs originated from plant molecules or their derivatives. Yet only a small fraction of plant chemical diversity has been explored due to limitations of traditional bioassay-guided isolation and toxicity evaluation.

Conventional phytochemical drug discovery involves extraction, isolation, structure elucidation, biological testing and safety evaluation. This process is labor-intensive and often results in rediscovery of known compounds. Furthermore, toxicity assessment requires animal studies and long-term evaluation. Computational prediction of bioactivity and toxicity can prioritize promising molecules before experimental testing.

Artificial intelligence refers to computational systems capable of learning patterns from data and making predictions. Machine learning algorithms analyze chemical descriptors derived from molecular structures and correlate them with biological or toxicological endpoints. AI approaches have transformed pharmaceutical research, especially in virtual screening and predictive toxicology. Natural products research is increasingly adopting AI to explore complex phytochemical space.

AI-based phytochemical prediction integrates cheminformatics, pharmacology and data science. By learning structure–activity relationships, models can estimate therapeutic potential, target affinity and safety risks of plant compounds. Such predictions guide selection of candidates for further validation. This review summarizes AI methodologies for predicting phytochemical bioactivity and toxicity, available datasets, modeling workflows, applications and future prospects.

CHEMICAL DIVERSITY OF PHYTOCHEMICALS AND NEED FOR PREDICTION

Plant metabolites display extraordinary structural diversity compared to synthetic libraries. Unique scaffolds, stereochemistry, glycosylation and functional groups contribute to varied biological effects. This diversity is beneficial for drug discovery but challenging for systematic screening.

1. Diversity and Complexity

Phytochemicals vary widely in molecular weight, polarity, and functionalization. Examples include:

- Alkaloids with nitrogen heterocycles
- Flavonoids with polyphenolic rings
- Terpenoids with isoprene units
- Saponins with glycoside chains
- Phenolic acids and tannins

Such diversity makes experimental evaluation of all compounds impractical. AI methods can explore large chemical spaces rapidly.

2. Challenges in Experimental Screening

Traditional phytochemical screening faces several limitations:

- Low yield of pure compounds
- Complex mixtures in extracts
- Limited bioassay throughput
- Costly toxicity testing
- Redundant isolation of known molecules

Predictive AI models help prioritize novel and safe compounds, reducing unnecessary experiments.

ARTIFICIAL INTELLIGENCE APPROACHES IN PHYTOCHEMICAL BIOACTIVITY PREDICTION

AI methods learn relationships between chemical structure and biological effect. Various algorithms are used depending on dataset size and complexity.

1. Quantitative Structure–Activity Relationship (QSAR)

QSAR modeling correlates molecular descriptors with biological activity values. Descriptors represent structural features such as:

- Molecular weight
- LogP (lipophilicity)

- Topological indices
- Hydrogen bond donors/acceptors
- Electronic properties

Machine learning algorithms (linear regression, support vector machines, random forest) learn predictive equations linking descriptors to activity. For phytochemicals, QSAR models predict enzyme inhibition, receptor binding or antimicrobial potency.

Advantages:

- Interpretable relationships
- Works with moderate datasets
- Useful for activity optimization

Limitations:

- Requires accurate activity data
- Limited extrapolation to novel scaffolds

2. Machine Learning Classification Models

Classification models categorize phytochemicals as active/inactive or toxic/non-toxic. Algorithms include:

- Decision trees
- Random forest
- k-nearest neighbors
- Naive Bayes
- Gradient boosting

These models handle nonlinear relationships and complex descriptor interactions. Random forest is particularly popular due to robustness and ability to handle noisy phytochemical data.

Applications include predicting:

- Anticancer activity
- Antimicrobial activity

- Anti-inflammatory potential
- Enzyme inhibition

3. Deep Learning Approaches

Deep learning uses artificial neural networks with multiple layers to learn hierarchical molecular features. Representations may include:

- Molecular fingerprints
- Graph neural networks (GNN)
- SMILES strings encoding

Deep learning models capture subtle structural patterns beyond handcrafted descriptors. They are suitable for large phytochemical datasets and multi-target prediction.

Examples of deep learning applications:

- Predicting phytochemical–protein interactions
- Multitask activity prediction
- Structure-based toxicity estimation
- Generating novel phytochemical analogs

However, deep learning requires large curated datasets and high computational resources.

4. Molecular Docking Combined with AI

Docking predicts binding of phytochemicals to biological targets. AI integration improves scoring and ranking of docking poses. Workflow:

- Dock phytochemicals to target protein
- Extract interaction features
- Train ML model on known ligand–target data
- Predict binding affinity

This hybrid approach enhances reliability compared to docking alone. It is widely used for enzyme inhibitors and receptor ligands from plants.

5. Network Pharmacology and Systems AI

Many phytochemicals act on multiple targets. Network pharmacology analyzes compound–target–pathway interactions. AI methods identify:

- Multi-target profiles
- Synergistic phytochemical combinations
- Disease pathway modulation

This is particularly relevant for herbal medicines where mixtures produce therapeutic effects.

AI-BASED TOXICITY PREDICTION OF PHYTOCHEMICALS

Safety evaluation is essential for natural product drug discovery. Some plant compounds exhibit hepatotoxicity, mutagenicity or cardiotoxicity. AI predictive toxicology reduces reliance on animal testing.

1. Toxicological Endpoints Predicted by AI

Common toxicity endpoints modeled include:

- Acute toxicity (LD50)
- Hepatotoxicity
- Mutagenicity
- Carcinogenicity
- Cardiotoxicity
- Skin sensitization
- Developmental toxicity

Models classify compounds as toxic/non-toxic or estimate toxicity values.

2. Structural Alerts and Rule-Based Systems

Certain chemical motifs are associated with toxicity (e.g., nitro groups, epoxides, quinones). Rule-based AI systems detect such structural alerts in phytochemicals. These provide quick hazard screening before advanced modeling.

3. Machine Learning Toxicity Models

Descriptors of phytochemicals are used to train toxicity prediction models. Algorithms include:

- Random forest toxicity classifiers

- Support vector regression for LD50
- Deep neural networks for multi-endpoint toxicity

Such models identify potentially harmful plant compounds early in discovery.

4. ADMET Prediction

AI tools estimate pharmacokinetic and safety properties collectively termed ADMET:

- Absorption
- Distribution
- Metabolism
- Excretion
- Toxicity

For phytochemicals, ADMET prediction evaluates:

- Oral bioavailability
- Blood–brain barrier penetration
- Cytochrome enzyme interaction
- Hepatotoxicity risk

Early ADMET screening avoids later clinical failures.

DATABASES AND DATASETS FOR PHYTOCHEMICAL AI MODELING

AI models require high-quality chemical and biological data. Several phytochemical and bioactivity databases support modeling.

Table 1: Key Data Sources for Phytochemical AI Prediction

Data Type	Content	Use in AI Modeling
Phytochemical structures	Molecular structures from plants	Descriptor generation
Bioactivity assays	Enzyme/receptor activity data	QSAR training
Toxicity records	Toxicological endpoints	Toxicity prediction
Target proteins	Protein structures	Docking and ML
Herbal formulations	Multi-compound data	Network pharmacology

Data curation is critical because errors propagate into AI predictions. Standardization of chemical structures and activity units is necessary.

WORKFLOW OF AI-BASED PHYTOCHEMICAL PREDICTION

A typical AI modeling pipeline includes several stages.

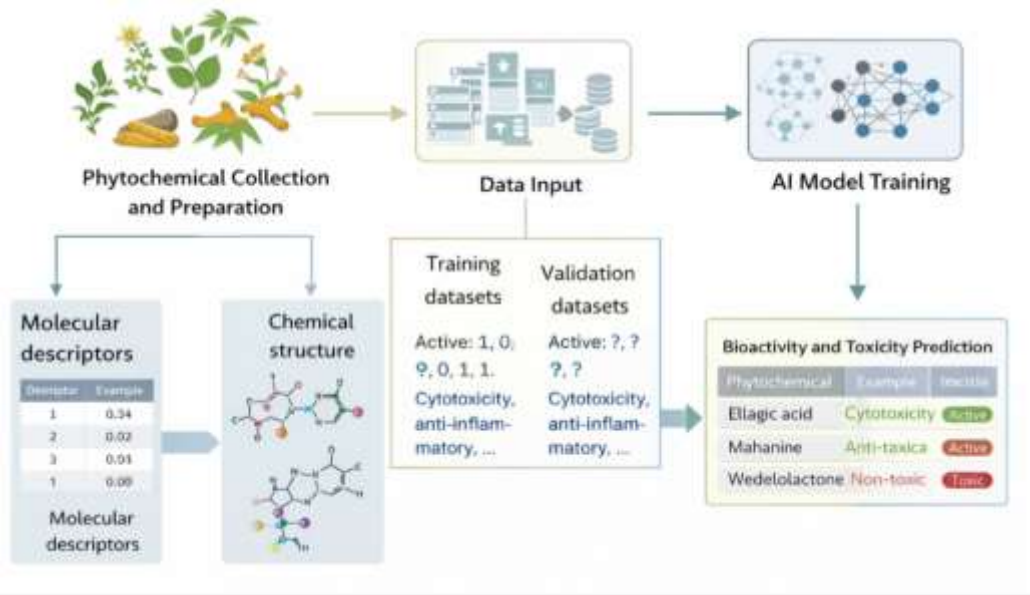

Figure 1: General workflow of AI prediction of phytochemical bioactivity and toxicity

1. Collection of phytochemical structures
2. Descriptor or fingerprint generation
3. Dataset preparation (training/test split)
4. Model selection and training
5. Validation and performance evaluation
6. Prediction of new phytochemicals
7. Experimental confirmation

1. Molecular Descriptor Generation

Chemical structures are converted into numerical descriptors representing:

- Physicochemical properties
- Topology
- Geometry
- Electronic features

Fingerprints encode presence of substructures. Descriptor selection affects model performance.

2. Model Training and Validation

Data is divided into training and validation sets. Algorithms learn patterns from training data. Validation metrics include:

- Accuracy
- Sensitivity
- Specificity
- ROC-AUC
- Mean squared error

Cross-validation ensures model generalization.

3. Applicability Domain

AI predictions are reliable only for compounds similar to training data. Applicability domain analysis ensures phytochemicals fall within model chemical space. This is crucial due to unique natural product structures.

APPLICATIONS IN PHYTOCHEMICAL DRUG DISCOVERY

AI prediction has enabled discovery of bioactive phytochemicals in several therapeutic areas.

1. Anticancer Phytochemicals

Machine learning models trained on anticancer compound datasets identify plant molecules with cytotoxic potential. Predicted activities include:

- Kinase inhibition
- DNA intercalation
- Apoptosis induction
- Anti-angiogenic effects

AI prioritizes candidates for cancer cell testing, reducing experimental workload.

2. Antimicrobial Phytochemicals

AI models predict antibacterial and antifungal activity of plant metabolites. Features linked to membrane disruption or enzyme inhibition are identified. This supports search for alternatives to synthetic antibiotics.

3. Anti-Inflammatory and Antioxidant Compounds

Flavonoids and phenolics often show anti-inflammatory effects. AI models predict inhibition of inflammatory enzymes or cytokine pathways. Antioxidant activity prediction helps identify compounds with radical scavenging potential.

4. Neuroprotective Phytochemicals

AI docking and ML predict phytochemicals interacting with neurological targets such as cholinesterase or amyloid aggregation. This aids discovery of plant-based neuroprotective agents.

CASE EXAMPLES OF AI PREDICTION IN PHYTOCHEMICALS

Several studies demonstrated successful AI-guided phytochemical discovery.

- Machine learning identified flavonoid derivatives with predicted anticancer activity later confirmed experimentally.
- QSAR models predicted alkaloid antimicrobial potency consistent with laboratory results.
- Toxicity models flagged hepatotoxic plant compounds matching known toxic herbs.

These examples show AI can guide experimental phytochemistry efficiently.

ADVANTAGES OF AI-BASED PREDICTION IN NATURAL PRODUCTS RESEARCH

AI offers multiple benefits compared to conventional screening.

Table 2: Advantages of AI Prediction in Phytochemical Research

Aspect	Traditional Screening	AI Prediction
Speed	Slow	Rapid
Cost	High	Lower
Animal testing	Required	Reduced
Chemical space	Limited	Large
Early toxicity detection	Late stage	Early stage
Multi-target analysis	Difficult	Feasible

AI therefore accelerates phytochemical drug discovery while improving safety assessment.

CHALLENGES AND LIMITATIONS

Despite promise, AI prediction of phytochemical bioactivity and toxicity faces challenges.

1. **Limited and Biased Data**

   Phytochemical datasets are smaller than synthetic drug databases. Many plant compounds lack bioactivity data. Bias toward well-studied compounds affects model generalization.

2. **Structural Novelty**

   Natural products often have rare scaffolds not present in training sets. AI models may extrapolate poorly outside known chemical space.

3. **Data Quality Issues**

   Experimental variability, inconsistent units and incorrect structures reduce model reliability.

   Careful curation is required.

4. **Interpretability of AI Models**

   Deep learning models act as “black boxes”. Understanding why a phytochemical is predicted active or toxic is difficult. Explainable AI methods are needed.

5. **Validation Gap**

   Predicted activities require experimental confirmation. Lack of validation limits practical adoption in phytochemistry labs.

EMERGING TRENDS AND FUTURE PERSPECTIVES

AI prediction of phytochemical bioactivity and toxicity continues to evolve.

1. **Explainable AI**

Interpretability techniques identify structural features responsible for activity or toxicity. This supports rational phytochemical optimization.

2. **Multi-Omics Integration**

Combining genomics, metabolomics and transcriptomics with AI reveals plant biosynthetic pathways and biological targets.

3. **Generative AI for Natural Products**

AI models can design novel phytochemical-like molecules with predicted activity and safety. This expands natural product chemical space.

4. **Hybrid Experimental–AI Pipelines**

Integration of AI prediction with high-throughput bioassays enables iterative discovery cycles: Prediction → Testing → Model refinement → New prediction

5. Personalized Herbal Medicine

AI may predict individual response or toxicity to phytochemicals based on genetic and metabolic profiles.

CONCLUSION

AI-based prediction of phytochemical bioactivity and toxicity represents a transformative approach in natural product research. Machine learning, QSAR, deep learning and network pharmacology enable rapid screening of plant metabolites for therapeutic potential and safety risks. These methods reduce experimental burden, cost and animal testing while exploring vast phytochemical diversity. Applications span anticancer, antimicrobial, anti-inflammatory and neuroprotective phytochemical discovery. However, challenges such as limited datasets, structural novelty and interpretability remain. Future advances in explainable AI, multi-omics integration and hybrid experimental–computational workflows will enhance reliability and adoption. AI-guided phytochemistry has strong potential to accelerate discovery of safe and effective plant-derived drugs.

REFERENCES

1. Chen Y., Garcia de Lomana M., Friedrich N.O., Kirchmair J. (2018). Characterization of the chemical space of known and readily obtainable natural products. *J Chem Inf Model*, 58, 1518–1532.
2. Lo Y.C., Rensi S.E., Torng W., Altman R.B. (2018). Machine learning in chemoinformatics and drug discovery. *Drug Discov Today*, 23, 1538–1546.
3. Sliwoski G., Kothiwale S., Meiler J., Lowe E.W. (2014). Computational methods in drug discovery. *Pharmacol Rev*, 66, 334–395.
4. Mayr A., Klambauer G., Unterthiner T., Hochreiter S. (2016). Deep learning in drug discovery. *Mol Inform*, 35, 95–107.
5. Ekins S. (2016). The next era: deep learning in pharmaceutical research. *Pharm Res*, 33, 259–263.
6. Fourches D., Muratov E., Tropsha A. (2015). Trust, but verify: data curation in QSAR modeling. *J Chem Inf Model*, 55, 1121–1131.
7. Raies A.B., Bajic V.B. (2016). In silico toxicology: computational methods for toxicity prediction. *Wiley Interdiscip Rev Comput Mol Sci*, 6, 147–172.
8. Wink M. (2015). Modes of action of herbal medicines and plant secondary metabolites.

Medicines, 2, 251–286.

9. Vamathevan J. et al. (2019). Applications of machine learning in drug discovery. *Nat Rev Drug Discov*, 18, 463–477.
10. Stokes J.M. et al. (2020). Deep learning approach to antibiotic discovery. *Cell*, 180, 688–702.

Cite as:

Dr. Farhad Rahman, Shahnaj Anjum (2026). AI-Based Prediction of Phytochemical Bioactivity and Toxicity. International Journal of Pharmacognosy and Phytochemical Sciences, 4(1), 7-19.

https://doi.org/10.5281/zenodo.19707551