Explainable Molecular AI for Phytopharmacology: Linking Chemical Features, Target Interactions, Pathway Effects, and Therapeutic Decision Confidence

David Osei[1]*, Akua Afriyie[1], Kofi Adu[2]

[1]Department of AI for Traditional Medicine and Pharmacokinetics, Faculty of Pharmacy, University of Ghana, Accra, Ghana.

[2]Department of Natural Product Informatics, Faculty of Pharmacy, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

ABSTRACT

Phytopharmacology increasingly uses molecular artificial intelligence to interpret plant-derived chemical diversity, predict bioactivity, infer target interactions, and prioritize candidates for experimental follow-up. However, predictive performance alone is insufficient when computational outputs are used to support therapeutic interpretation, mechanism hypotheses, safety reasoning, or candidate selection. This article proposes an explainable molecular AI framework for phytopharmacology that links chemical feature explanations, target-level interpretation, pathway-level reasoning, ADMET and toxicity interpretation, uncertainty assessment, expert review, and therapeutic decision confidence. The framework treats explainability as a structured decision-support process rather than as proof of mechanism, safety, or efficacy. At the chemical level, explanations may connect molecular descriptors, fingerprints, graph features, substructures, attention patterns, and counterfactual changes to predicted bioactivity. At the target level, interpretable models may support hypotheses about phytochemical–target interactions, target-family relevance, and binding plausibility, while requiring external validation. At the pathway level, network pharmacology and systems-level interpretation can organize predicted targets into biological hypotheses, but remain vulnerable to database bias, overconnected nodes, and weak causal specificity. Decision confidence is therefore framed as a convergence judgment across explanation quality, uncertainty, applicability domain, data provenance, pharmacological plausibility, ADMET concerns, and expert review. The main contribution is an integrated architecture for transparent candidate prioritization in natural product drug discovery. The framework can help researchers avoid overreliance on opaque predictions, visually persuasive explanations, isolated target claims, or unsupported pathway diagrams, while guiding reproducible reporting, validation planning, and future development of responsible explainable AI tools for phytopharmacology.

Key Words: *Explainable artificial intelligence, Molecular AI, Phytopharmacology, Chemical feature attribution, Target prediction, Pathway interpretability*

eIJPPR 2025; 14(2):70-79

HOW TO CITE THIS ARTICLE: Osei D, Afriyie A, Adu K. Explainable Molecular AI for Phytopharmacology: Linking Chemical Features, Target Interactions, Pathway Effects, and Therapeutic Decision Confidence. Int J Pharm Phytopharmacol Res. 2025;15(2):70-9. https://doi.org/10.51847/1xwhZbanL7

INTRODUCTION

Phytopharmacology is increasingly positioned at the intersection of natural product chemistry, molecular pharmacology, computational biology, and translational drug discovery. Plant-derived molecules and other natural products remain important sources of pharmacologically relevant chemical matter, which makes transparent

Corresponding author: David Osei
Address: Department of AI for Traditional Medicine and Pharmacokinetics, Faculty of Pharmacy, University of Ghana, Accra, Ghana.
E-mail: ✉ david.osei@ug.edu.gh
Received: 08 January 2025; **Revised:** 18 April 2025; **Accepted:** 23 April 2025

approaches to candidate prioritization especially valuable when large chemical spaces must be interpreted before experimental follow-up [1]. The challenge is not simply to screen more compounds, but to explain why particular phytochemicals, molecular features, predicted targets, or mechanistic hypotheses should be considered biologically plausible.

Deep learning has expanded the ability of computational systems to model molecular properties, bioactivity patterns, and drug-discovery tasks that depend on complex chemical and biological representations [2]. In phytopharmacology, this creates opportunities to move beyond descriptive cataloguing of plant-derived molecules toward model-assisted interpretation of bioactivity, target interaction, and safety-related properties. Yet the same complexity that makes deep models attractive can also obscure the chemical reasoning behind predictions, particularly when models rely on learned embeddings, molecular graphs, sequence encodings, or high-dimensional fingerprints.

Machine learning has also become relevant across discovery and development workflows, including target identification, lead optimization, property prediction, safety assessment, and translational decision support [3]. For phytochemical research, these tasks are complicated by structural diversity, incomplete annotation, variable assay quality, polypharmacology, and context-dependent biological effects. A framework for explainable molecular AI must therefore connect computational prediction with pharmacological interpretation while acknowledging that model outputs require experimental validation and expert judgment.

Explainable AI has been proposed as a way to improve transparency in drug discovery by making model predictions more interpretable to chemists, pharmacologists, and decision-makers [4]. However, explanation should not be treated as equivalent to biological truth. The aim of this article is to define an explainable molecular AI framework for phytopharmacology that links chemical features, target interactions, pathway effects, ADMET and toxicity interpretation, uncertainty, expert review, and therapeutic decision confidence. The central questions are how molecular explanations can support phytochemical interpretation, how target and pathway hypotheses can be evaluated responsibly, and how decision confidence can be assigned without overstating therapeutic relevance.

Explainable molecular AI in phytopharmacology

Explainable molecular AI in phytopharmacology can be defined as the use of interpretable or explanation-generating computational models to connect phytochemical representations with predicted biological, pharmacological, or safety-related outcomes in a form that supports expert evaluation. Molecular explanations should be assessed for fidelity to the model, stability across perturbations, chemical plausibility, and relevance to the task being interpreted [5]. This definition separates explanation quality from visual appeal: an attractive heat map, highlighted fragment, or attention pattern is useful only when it faithfully reflects the model's reasoning and can be interpreted within a chemically and biologically meaningful context.

In classical and modern QSAR settings, interpretability often depends on whether molecular descriptors, structural alerts, fragments, or physicochemical features can be connected to model outputs in a way that supports cautious chemical reasoning [6]. For phytopharmacology, descriptor-level interpretation can be useful when evaluating plant-derived molecules with known structural motifs, but it becomes weaker when descriptors are treated as direct biological mechanisms. A high-value explanation should clarify the relationship between molecular representation and prediction while preserving the distinction between statistical association, chemical plausibility, and experimentally demonstrated pharmacology.

Counterfactual molecular explanations provide another route to interpretability by asking what minimal structural or representational changes would alter a model prediction [7]. In phytochemical candidate prioritization, counterfactual reasoning may help researchers identify model-sensitive substructures, substituent patterns, or molecular features that influence predicted activity or safety. Nevertheless, counterfactuals remain model-dependent explanations rather than experimentally verified medicinal chemistry proposals. Their use in phytopharmacology should therefore be coupled with synthetic feasibility, natural-product plausibility, pharmacological relevance, and expert review.

Reliable explainable molecular AI also requires safeguards that are familiar from QSAR practice, including applicability-domain assessment, reproducibility checks, dataset transparency, and endpoint-specific validation [8]. These safeguards are especially important for plant-derived molecules because natural product structures may differ from the chemical spaces used to train many predictive models. A phytochemical explanation may be technically coherent within a model but still unreliable if the compound is outside the model's domain, if the endpoint is poorly annotated, or if the biological assay context does not match the intended interpretation.

Chemical features and target-level explanations

Chemical-level explanations form the first interpretive layer of the proposed framework because most molecular

AI predictions begin with representations of structure. Attribution methods can help identify chemical features associated with predicted binding behaviour in neural chemistry models, supporting a more explicit link between molecular patterns and target-related predictions [9]. In phytopharmacology, this layer may involve descriptors, fingerprints, substructure attribution, graph-node importance, fragment masking, saliency-like methods, or counterfactual changes that help explain why a model prioritizes a plant-derived molecule for a particular endpoint.

Graph neural networks are especially relevant because molecular graphs represent atoms, bonds, neighborhoods, and learned relational patterns, but their explanations require systematic evaluation rather than acceptance at face value. Quantitative evaluation of explainable graph neural networks for molecular property prediction shows that explanation methods can differ in reliability and should be tested for whether they actually capture model-relevant molecular information [10]. For phytochemical interpretation, this means that graph explanations should not be used merely to decorate predictions; they should be examined for stability, chemical coherence, and endpoint relevance.

Substructure-aware explanation provides a practical bridge between learned molecular representations and chemist-readable interpretation. Chemistry-intuitive graph explanations based on substructure masking can help identify fragments or molecular regions that influence model predictions in a way that is more interpretable than raw latent features [11]. In phytopharmacology, this is useful because many plant-derived molecules contain recognizable scaffolds, glycosides, phenolic motifs, alkaloid cores, terpenoid frameworks, or other structural patterns. Even when such explanations are chemically meaningful, they should be framed as model-derived hypotheses rather than direct evidence of target engagement or therapeutic effect.

Target-level explanations extend chemical interpretation toward biological plausibility by asking how a phytochemical representation relates to a predicted protein, target family, or interaction context. Molecular interaction transformer approaches illustrate how compound and protein representations can be integrated for drug–target interaction prediction [12]. For an explainable phytopharmacology workflow, target-level interpretation should connect predicted interactions with chemical-feature evidence, protein-family relevance, available pharmacology, assay feasibility, and uncertainty. A predicted target should be treated as a hypothesis for validation, not as confirmation that a compound modulates that target in a biological system.

Table 1 compares chemical-feature and target-level explanation strategies for molecular AI in phytopharmacology, highlighting their interpretation value, validation needs, and limitations.

Table 1. Chemical-Feature and Target-Level Explanation Strategies in Explainable Molecular AI for Phytopharmacology: Representation Inputs, Explanation Outputs, Biological Interpretation, Validation Needs, Limitations, and Decision-Relevance

Explanation strategy	Molecular or biological input	AI model context	Explanation output	Phytochemical interpretation value	Target-level relevance	Validation requirement	Interpretation limitation	Decision-confidence contribution
Descriptor interpretation	Physicochemical descriptors, topological descriptors, constitutional descriptors	QSAR, classical machine learning, interpretable regression or tree-based models	Descriptor contribution or direction of influence	Helps connect prediction to recognizable chemical properties such as polarity, size, lipophilicity, or hydrogen-bonding capacity	Indirect; may suggest properties compatible with target binding or permeability	Applicability-domain assessment, endpoint-specific validation, expert chemical review	Descriptor effects may be statistical rather than mechanistic	Supports moderate confidence when descriptor effects are stable, plausible, and within domain
Fingerprint attribution	Molecular fingerprints, fragment keys, circular fingerprints	Bioactivity prediction, similarity-based learning, ensemble models	Important fingerprint bits or fragment-like signals	Highlights structural patterns associated with predicted activity	Indirect to moderate; may support target-family hypotheses when fragments align with known pharmacophores	Fragment mapping, similarity analysis, bioactivity assay confirmation	Fingerprint bits may be difficult to map uniquely to interpretable chemistry	Supports candidate ranking when attribution is chemically interpretable and reproducible

Substructure attribution	Molecular fragments, scaffolds, functional groups, graph neighborhoods	Graph neural networks, fragment-based models, attribution models	Highlighted substructures or masked fragments influencing prediction	Useful for interpreting phytochemical scaffolds and substituent patterns	Moderate; may suggest substructures relevant to binding hypotheses	Fragment perturbation, model-fidelity checks, medicinal chemistry review	Highlighted substructures do not prove biological causality	Increases confidence when substructures align with known chemistry and target plausibility
Graph explanation	Atom-bond graphs, node features, edge features, molecular neighborhoods	Graph neural networks for property or activity prediction	Atom, bond, edge, or subgraph importance	Captures local and relational molecular features beyond fixed descriptors	Moderate; can support target-level hypotheses when linked to interaction models	Explanation benchmarking, stability analysis, task-specific evaluation	Graph explanations may vary across methods and perturbations	Supports confidence only when quantitatively evaluated and chemically coherent
Attention interpretation	SMILES tokens, molecular sequences, protein sequences, transformer embeddings	Transformer-based molecular or drug–target interaction models	Attention patterns or token-level emphasis	May show model focus across molecular tokens or compound–protein representations	Moderate to high in DTI contexts, but only when interpreted cautiously	Attention sanity checks, comparison with attribution methods, target assay validation	Attention weights are not automatically explanations or biological evidence	Provides supportive context, not standalone confidence
Counterfactual explanation	Molecular structure, latent representation, proposed structural change	Model-agnostic or model-specific counterfactual explanation systems	Minimal molecular changes predicted to alter output	Helps identify model-sensitive structural features and possible optimization directions	Indirect; may suggest changes affecting predicted target interaction	Chemical feasibility review, synthetic plausibility, experimental testing	Counterfactual changes may be unrealistic, unstable, or outside natural-product chemistry	Supports decision refinement when chemically feasible and uncertainty-aware
Target-interaction explanation	Compound representation, protein sequence, target-family data, interaction prediction	Drug–target interaction prediction, target inference, multitask bioactivity models	Predicted target relation, target-family rationale, interaction hypothesis	Connects phytochemical features to possible pharmacological targets	Direct; supports target prioritization and assay planning	Target engagement assays, orthogonal bioactivity tests, biological-context review	Target prediction is not target validation	Improves confidence when target prediction converges with chemical, pathway, and assay evidence
Binding-plausibility reasoning	Predicted interaction context, known target class, chemical compatibility	DTI models, structure-aware models where available, expert review	Plausibility judgment rather than definitive binding proof	Helps screen whether a phytochemical hypothesis is chemically reasonable	Direct but hypothesis-level only	Biophysical or biochemical target engagement validation	Plausibility can be biased by prior expectations or incomplete target data	Supports prioritization for validation rather than therapeutic claims
Expert chemical review	Model explanation outputs plus chemical knowledge	Human-in-the-loop interpretation	Expert judgment on explanation plausibility and feasibility	Prevents overinterpretation of visually plausible but chemically weak explanations	Supports target-hypothesis triage	Independent review, documented rationale, reproducible reporting	Expert judgment can be subjective if not structured	Strengthens confidence when combined with uncertainty and validation planning

ISSN (Online) 2249-6084 (Print) 2250-1029

Pathway-level interpretability

Pathway-level interpretability is the bridge between molecular prediction and biological mechanism in explainable phytopharmacology. Plant-derived molecules frequently act through multi-target and context-dependent pharmacological patterns, so target interpretation should not stop at a ranked list of predicted proteins. In silico polypharmacology provides a rationale for organizing natural-product activity across multiple targets, target families, and network contexts rather than forcing phytochemicals into single-target assumptions [13]. In the proposed framework, pathway interpretation begins only after chemical-feature and target-level explanations have been assessed for plausibility.

Network pharmacology can help convert target hypotheses into systems-level interpretations by examining how predicted targets may relate to disease networks, biological processes, and causal mechanism hypotheses [14]. However, pathway-level interpretability should be cautious: a network diagram does not prove causality, a connected target does not prove biological relevance, and a pathway label does not confirm therapeutic mechanism. The value of pathway interpretation lies in structuring hypotheses, identifying mechanistic coherence, and prioritizing experimental tests, not in replacing biological validation.

Functional enrichment and visualization methods can support this process by translating target lists or gene sets into interpretable biological functions, pathways, and process-level summaries [15]. In phytopharmacology, this can help researchers ask whether predicted targets converge on plausible inflammatory, metabolic, oxidative-stress, immune, neuronal, microbial, or cancer-related pathways. Yet enrichment results are sensitive to input target quality, database coverage, pathway redundancy, annotation bias, and overconnected genes. Therefore, pathway-level explanations should include uncertainty about the target list, the pathway database, and the biological context in which the pathway is being interpreted.

Mechanism-of-action inference requires integration across chemical, target, pathway, phenotypic, omics, and validation evidence rather than reliance on a single computational explanation layer [16]. A responsible explainable molecular AI workflow should therefore distinguish predicted target interaction from target engagement, pathway mapping from pathway modulation, and mechanism hypothesis from mechanism confirmation. Pathway interpretability becomes most useful when it identifies testable mechanistic questions: which targets require biochemical validation, which pathways require cellular confirmation, which biomarkers should be measured, and which safety liabilities should be evaluated before therapeutic claims are made.

Figure 1 illustrates the multi-level explanation pathway linking phytochemical features, predicted targets, pathway effects, mechanism interpretation, and therapeutic decision confidence.

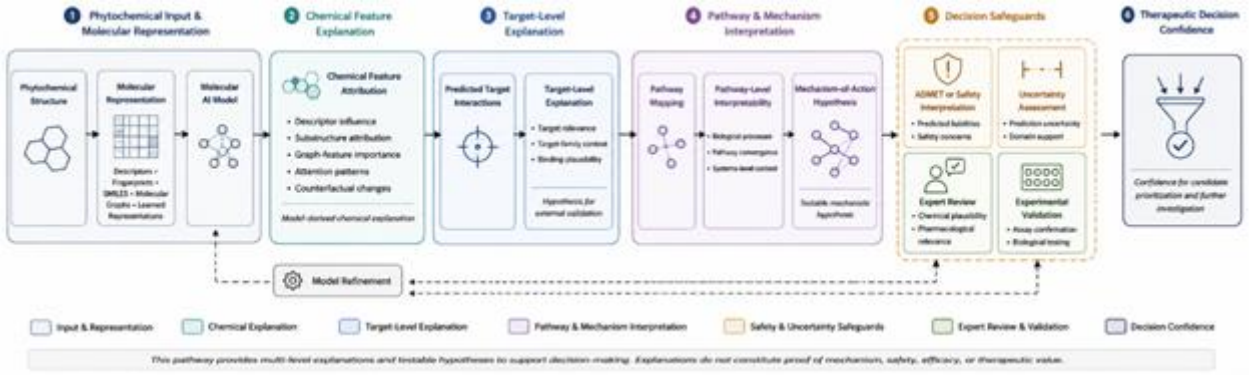

Figure 1. Multi-Level Explanation Pathway for Explainable Molecular AI in Phytopharmacology

Alt-text description

A multi-level explanation diagram showing phytochemical chemical features feeding into molecular representations and AI models. The model outputs are interpreted through chemical feature attribution, target interaction prediction, pathway mapping, mechanism-of-action inference, uncertainty assessment, and expert review. These layers converge into therapeutic decision confidence, with feedback loops from validation back to model refinement.

Decision confidence and expert review

Decision confidence in explainable molecular AI should be interpreted as a structured judgment about whether a phytochemical candidate deserves further investigation, not as a guarantee of efficacy, safety, or clinical value. Uncertainty quantification is therefore essential because molecular prediction models can produce plausible outputs even when the prediction is poorly supported by the training domain, endpoint quality, or chemical context

[17]. In phytopharmacology, uncertainty should be attached to bioactivity predictions, target hypotheses, ADMET estimates, toxicity alerts, and pathway interpretations.

Evidential deep learning provides one approach for linking molecular predictions with uncertainty-aware prioritization, especially when models are used to guide discovery under incomplete data conditions [18]. In the proposed framework, uncertainty contributes to decision confidence by identifying which predictions are robust enough to support prioritization and which should be downgraded, repeated, or experimentally checked before interpretation. A phytochemical with a plausible explanation but high uncertainty should not be advanced as confidently as one with convergent chemical, target, pathway, safety, and uncertainty evidence.

Scalable uncertainty estimation is also important when large phytochemical libraries or natural-product databases are screened computationally [19]. Library-scale prediction can create a false sense of precision if every candidate receives a ranked output without a clear indication of model confidence, applicability domain, and data quality. For decision support, uncertainty should be combined with explanation stability, domain similarity, endpoint reliability, and expert review so that researchers can separate high-priority validation candidates from weakly supported computational artefacts.

Expert review is the final interpretive safeguard before computational explanations influence therapeutic claims or experimental resource allocation. Critical appraisal of AI in drug discovery emphasizes that model performance and apparent interpretability must be separated from realistic translational impact [20]. In phytopharmacology, experts should ask whether the highlighted chemical features are pharmacologically plausible, whether predicted targets are assayable and biologically relevant, whether pathway interpretations are specific enough to guide experiments, whether safety signals are concerning, and whether the proposed decision is proportionate to the evidence.

Table 2 presents a decision-confidence and expert-review matrix for evaluating explainable molecular AI outputs in phytopharmacology.

Table 2. Decision-Confidence and Expert-Review Matrix for Explainable Molecular AI in Phytopharmacology: Evidence Layers, Confidence Signals, Uncertainty Sources, Expert Checks, Validation Requirements, Translational Risks, and Candidate Actions

Decision layer	Evidence source	Confidence signal	Uncertainty source	Expert-review question	Validation requirement	Translational risk if ignored	Interpretation safeguard	Candidate action
Chemical input quality	Curated phytochemical structure, source annotation, molecular standardization	Structure is well defined, non-duplicated, and chemically plausible	Ambiguous stereochemistry, missing metadata, inconsistent structure records	Is the compound identity sufficiently reliable for modeling?	Structure verification, provenance review, curation audit	False prioritization caused by incorrect or poorly curated input	Document data source, curation rules, and structural assumptions	Retain only if identity and representation are adequate
Molecular representation	Descriptors, fingerprints, SMILES, molecular graphs, learned embeddings	Representation captures relevant chemical features for the endpoint	Representation mismatch, tokenization artefacts, graph-feature limitations	Does the representation suit the phytochemical class and task?	Compare representation types and check domain coverage	Model may learn irrelevant or unstable patterns	Report representation choice and limitations	Compare models or downgrade if representation is weak
Chemical explanation	Feature attribution, substructure explanation, descriptor interpretation, counterfactual reasoning	Explanation is stable, chemically interpretable, and task-relevant	Explanation instability, poor fidelity, non-plausible highlighted features	Does the explanation make chemical sense?	Perturbation checks, expert chemical review	Visually appealing but misleading explanation	Require fidelity and plausibility checks	Advance only if explanation is coherent
Target-level hypothesis	Target prediction, DTI model output, target-family reasoning	Target aligns with chemical explanation and known pharmacological context	Out-of-domain target model, weak protein representation, limited assay evidence	Is the target biologically plausible and testable?	Biochemical or cellular target engagement assay	Unsupported target claims may drive false mechanism narratives	Treat target as a hypothesis, not validation	Prioritize for target confirmation if feasible

ISSN (Online) 2249-6084 (Print) 2250-1029

Pathway interpretation	Target clusters, network pharmacology, pathway mapping, mechanism hypothesis	Predicted targets converge on biologically coherent pathways	Database bias, pathway redundancy, overconnected targets	Is the pathway interpretation specific and causally plausible?	Pathway biomarker assays, omics validation, perturbation experiments	Unsupported pathway diagrams may imply false mechanism	Require context-specific pathway review	Use pathway results to design experiments
ADMET and toxicity review	ADMET prediction, toxicity model, structural alerts, safety explanation	Safety signals are interpretable and not contradictory	Endpoint uncertainty, alert incompleteness, model domain limits	Are there plausible liabilities requiring early testing?	ADMET assays, toxicity screens, pharmacokinetic evaluation	Unsafe candidates may be overprioritized	Separate predicted safety from demonstrated safety	Downgrade, redesign, or test safety concern
Uncertainty and applicability domain	Predictive uncertainty, calibration, domain similarity, model confidence	Low uncertainty and in-domain prediction across relevant layers	Dataset shift, sparse endpoint data, chemical novelty	Is the model confident for this compound and endpoint?	Calibration checks, external validation, domain analysis	Overconfident decisions from unsupported predictions	Report uncertainty with every decision layer	Advance only if uncertainty is acceptable
Expert review	Human-in-the-loop chemistry, pharmacology, toxicology, and assay feasibility assessment	Expert panel agrees that explanation is plausible and testable	Subjectivity, disciplinary bias, incomplete evidence	Would the explanation justify experimental resources?	Structured expert review and documented rationale	Poorly grounded computational claims may enter manuscripts or pipelines	Use predefined review questions	Approve, revise, downgrade, or reject candidate
Validation readiness	Assay feasibility, target engagement plan, mechanism tests, safety testing	Clear next-step experiments can test the hypothesis	Lack of assay access, weak endpoint match, unclear mechanism	What experiment would falsify or support the claim?	Orthogonal experimental validation	Computational prioritization may be mistaken for evidence	Link every claim to a validation plan	Move to validation only with explicit test plan

Proposed explainable AI framework

The proposed framework begins with curated phytochemical data and treats data provenance as the first explanation layer. Natural-product databases can provide structured access to plant-derived and other natural-product chemical entities, but computational interpretation depends on accurate structures, standardized representations, and transparent source metadata [21]. Before any model explanation is considered, the framework requires compound identity review, molecular standardization, duplicate handling, stereochemical awareness where possible, and documentation of the intended prediction endpoint.

The second layer converts phytochemical inputs into molecular representations suitable for interpretable modeling. These may include descriptors, fingerprints, SMILES strings, molecular graphs, learned embeddings, and multimodal combinations with biological data. The third layer applies explanation methods such as feature attribution, substructure attribution, graph explanations, attention interpretation, counterfactual explanations, uncertainty estimation, and human-in-the-loop review. Safety interpretation is integrated rather than appended at the end: machine learning toxicity prediction and structural-alert approaches can help identify interpretable safety concerns during candidate triage [22].

Figure 2 presents the proposed explainable molecular AI framework for integrating chemical explanations, target interactions, pathway effects, uncertainty, expert review, and therapeutic decision confidence.

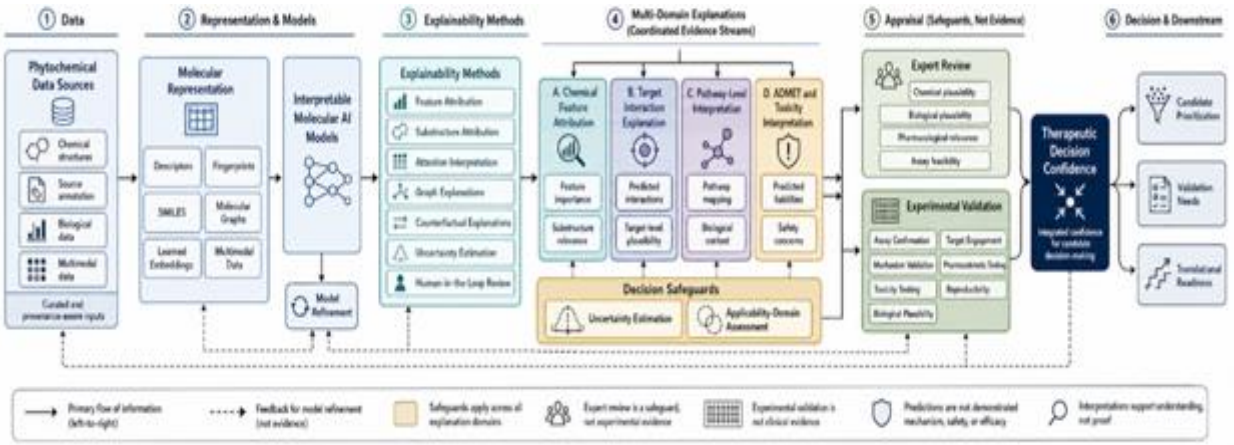

Figure 2. Explainable Molecular AI Framework for Therapeutic Decision Confidence in Phytopharmacology

Alt-text description

A conceptual framework showing phytochemical data entering an explainable molecular AI workflow. Molecular representations feed into interpretable models and explanation tools. Explanations are organized into chemical feature, target interaction, pathway effect, ADMET safety, uncertainty, and expert-review layers. These layers feed into a decision-confidence module that guides candidate prioritization, validation needs, and translational readiness.

The fourth layer organizes explanations into chemical-feature, target-interaction, pathway-effect, and safety-evidence modules. Toxicity-focused models with contrastive molecular explanations illustrate how prediction outputs can be paired with molecular-level interpretive cues that support expert safety review [23]. In the framework, such outputs are not used to declare a candidate safe. Instead, they identify safety questions, structural liabilities, endpoint uncertainty, and experimental tests that should be addressed before a phytochemical is advanced.

The fifth layer introduces expert review as a formal decision point rather than an informal afterthought. Human-centered concepts of explainability and causability emphasize that explanations must be understandable, usable, and meaningful to expert decision-makers, particularly when biomedical interpretation is involved [24]. In phytopharmacology, expert review should include medicinal chemistry, pharmacology, toxicology, computational biology, and assay-design perspectives where feasible. The expert panel should assess whether the model explanation is chemically plausible, whether the predicted target is biologically relevant, whether pathway interpretation is specific, and whether validation is practical.

The final layer converts integrated evidence into therapeutic decision confidence. This layer does not assign invented numerical confidence scores; instead, it uses structured categories such as high-priority validation candidate, conditional candidate requiring additional evidence, safety-concern candidate, out-of-domain candidate, or low-confidence computational hypothesis. Chemical and biological data quality strongly constrain what AI systems can realistically deliver in drug discovery, so the framework requires every decision to be accompanied by uncertainty, applicability-domain assessment, reproducibility information, and a validation plan [25]. Feedback loops from experimental validation and expert review return to data curation, representation choice, model selection, and explanation appraisal.

Practical implications

For phytopharmacology and natural product drug discovery, the framework encourages researchers to preserve the chemical diversity and discovery value of natural products rather than reducing candidates to opaque rank-ordered screening outputs. Retrospective analysis of natural-product discovery trends highlights the importance of natural-product chemical space for future discovery, which supports a practical need for AI workflows that remain interpretable, diversity-aware, and chemically grounded [26]. In practice, this means reporting the molecular representation, explanation method, model domain, target hypothesis, pathway interpretation, safety concern, and validation plan for each prioritized candidate. For ADMET and safety evaluation, the framework helps prevent premature claims based on isolated toxicity predictions or visually persuasive explanation graphics. Descriptor-based explainable toxicity modeling can support safety-focused candidate review when the endpoint, descriptors, and interpretation limits are clearly stated [27]. In phytopharmacology, such explanations may be useful for early triage, but they should be connected to pharmacokinetic testing, toxicity assays, metabolism considerations, and expert toxicological review before being used to support translational claims.

For large-scale computational screening, the framework supports practical triage by linking ADMET platforms, molecular prediction, uncertainty assessment, and expert review. Machine learning ADMET platforms can evaluate large chemical libraries and help prioritize compounds for follow-up, but their outputs should be interpreted alongside applicability-domain limits, uncertainty, assay feasibility, and biological plausibility [28]. This approach can help researchers avoid overreliance on opaque predictions, isolated target claims, unsupported pathway diagrams, or unvalidated confidence outputs while improving reproducibility and transparency in candidate selection.

CONCLUSION

Explainable molecular AI can strengthen phytopharmacology by connecting chemical features, target interactions, pathway effects, safety interpretation, expert review, and therapeutic decision confidence within a transparent framework. Its value lies in improving interpretation, prioritization, hypothesis generation, and validation planning, not in proving mechanism, efficacy, safety, or clinical utility. The proposed framework emphasizes that therapeutic relevance depends on biological plausibility, uncertainty assessment, applicability-domain review, reproducibility, experimental validation, target engagement confirmation, safety evaluation, and integration with pharmacology and medicinal chemistry. Used responsibly, explainable molecular AI can help transform phytochemical screening from opaque prediction toward structured, evidence-aware decision support.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Newman DJ, Cragg GM. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. J Nat Prod. 2020;83(3):770-803. doi:10.1021/acs.jnatprod.9b01285

[2] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-50. doi:10.1016/j.drudis.2018.01.039

[3] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-77. doi:10.1038/s41573-019-0024-5

[4] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84. doi:10.1038/s42256-020-00236-4

[5] Wellawatte GP, Gandhi HA, Seshadri A, White AD. A perspective on explanations of molecular prediction models. J Chem Theory Comput. 2023;19(8):2149-60. doi:10.1021/acs.jctc.2c01235

[6] Polishchuk P. Interpretation of quantitative structure-activity relationship models: Past, present, and future. J Chem Inf Model. 2017;57(11):2618-39. doi:10.1021/acs.jcim.7b00274

[7] Wellawatte GP, Seshadri A, White AD. Model agnostic generation of counterfactual explanations for molecules. Chem Sci. 2022;13(13):3697-705. doi:10.1039/D1SC05259D

[8] Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, Poroikov V, et al. QSAR without borders. Chem Soc Rev. 2020;49(11):3525-64. doi:10.1039/D0CS00098A

[9] McCloskey K, Taly A, Monti F, Brenner MP, Colwell LJ. Using attribution to decode binding mechanism in neural network models for chemistry. Proc Natl Acad Sci U S A. 2019;116(24):11624-9. doi:10.1073/pnas.1820657116

[10] Rao J, Zheng S, Lu Y, Yang Y. Quantitative evaluation of explainable graph neural networks for molecular property prediction. Patterns (N Y). 2022;3(12):100628. doi:10.1016/j.patter.2022.100628

[11] Wu Z, Wang J, Du H, Jiang D, Kang Y, Li D, et al. Chemistry-intuitive explanation of graph neural networks for molecular property prediction with substructure masking. Nat Commun. 2023;14(1):2585. doi:10.1038/s41467-023-38192-3

[12] Huang K, Xiao C, Glass LM, Sun J. MolTrans: Molecular interaction transformer for drug-target interaction prediction. Bioinformatics. 2021;37(6):830-6. doi:10.1093/bioinformatics/btaa880

[13] Fang J, Liu C, Wang Q, Lin P, Cheng F. In silico polypharmacology of natural products. Brief Bioinform. 2018;19(6):1153-71. doi:10.1093/bib/bbx045

[14] Nogales C, Mamdouh ZM, List M, Kiel C, Casas AI, Schmidt HHHW. Network pharmacology: Curing causal mechanisms instead of treating symptoms. Trends Pharmacol Sci. 2022;43(2):136-50. doi:10.1016/j.tips.2021.11.004

[15] Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, et al. KOBAS-i: Intelligent prioritization and exploratory

visualization of biological functions for gene enrichment analysis. Nucleic Acids Res. 2021;49(W1):W317-25. doi:10.1093/nar/gkab447

[16] Trapotsi MA, Hosseini-Gerami L, Bender A. Computational analyses of mechanism of action (MoA): Data, methods and integration. RSC Chem Biol. 2022;3(2):170-200. doi:10.1039/D1CB00069A

[17] Hirschfeld L, Swanson K, Yang K, Barzilay R, Coley CW. Uncertainty quantification using neural networks for molecular property prediction. J Chem Inf Model. 2020;60(8):3770-80. doi:10.1021/acs.jcim.0c00502

[18] Soleimany AP, Amini A, Goldman S, Rus D, Bhatia SN, Coley CW. Evidential deep learning for guided molecular property prediction and discovery. ACS Cent Sci. 2021;7(8):1356-67. doi:10.1021/acscentsci.1c00546

[19] Scalia G, Grambow CA, Pernici B, Li YP, Green WH. Evaluating scalable uncertainty estimation methods for deep learning-based molecular property prediction. J Chem Inf Model. 2020;60(6):2697-717. doi:10.1021/acs.jcim.9b00975

[20] Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 1: Ways to make an impact, and why we are not there yet. Drug Discov Today. 2021;26(2):511-24. doi:10.1016/j.drudis.2020.12.009

[21] Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: Collection of Open Natural Products database. J Cheminform. 2021;13(1):2. doi:10.1186/s13321-020-00478-9

[22] Yang H, Sun L, Li W, Liu G, Tang Y. In silico prediction of chemical toxicity for drug design using machine learning methods and structural alerts. Front Chem. 2018;6:30. doi:10.3389/fchem.2018.00030

[23] Sharma B, Chenthamarakshan V, Dhurandhar A, Pereira S, Hendler JA, Dordick JS, et al. Accurate clinical toxicity prediction using multi-task deep neural nets and contrastive molecular explanations. Sci Rep. 2023;13(1):4908. doi:10.1038/s41598-023-31169-8

[24] Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov. 2019;9(4):e1312. doi:10.1002/widm.1312

[25] Bender A, Cortés-Ciriano I. Artificial intelligence in drug discovery: What is realistic, what are illusions? Part 2: A discussion of chemical and biological data. Drug Discov Today. 2021;26(4):1040-52. doi:10.1016/j.drudis.2020.11.037

[26] Pye CR, Bertin MJ, Lokey RS, Gerwick WH, Linington RG. Retrospective analysis of natural products provides insights for future discovery trends. Proc Natl Acad Sci U S A. 2017;114(22):5601-6. doi:10.1073/pnas.1614680114

[27] Jaganathan K, Tayara H, Chong KT. An explainable supervised machine learning model for predicting respiratory toxicity of chemicals using optimal molecular descriptors. Pharmaceutics. 2022;14(4):832. doi:10.3390/pharmaceutics14040832

[28] Swanson K, Walther P, Leitz J, Mukherjee S, Wu JC, Shivnaraine RV, et al. ADMET-AI: A machine learning ADMET platform for evaluation of large-scale chemical libraries. Bioinformatics. 2024;40(7):btae416. doi:10.1093/bioinformatics/btae416