Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 16 March 2026
Revised 20 April 2026
Accepted 10 May 2026

Natural Resources for Human Health 2026, Vol. 6 (3): 271–282

KEYWORDS:
Artificial intelligence; Machine learning; Deep learning; Natural products; Medicinal plants; Pharmacological activity prediction; Drug discovery.

AI-Based Prediction of Pharmacological Activities of Plant-Derived Natural Products: Advances, Applications, Challenges, and Future Perspectives

Balkrishna K. Patil[1], Swati Babaso Udugade[2*], Hemant Kumar[3], Dr Avijit Mazumder[4], Himanshu Makhija[5], Dr. Soundararajan K[6], Dr. D. Alex Anand[7], Natalya Voronina[8]

[1] Department of Computer Science and Engineering, Symbiosis Institute of Technology, Pune Campus, Symbiosis International (Deemed University), Pune, India.balkrishna.patil@sitpune.edu.in

[2] Krishna Institute of Pharmacy, Krishna Vishwa Vidyapeeth, "Deemed to be University", Taluka-Karad, Dist-Satara, Pin-415 539, Maharashtra, India. swatiudugade@gmail.com

[3]Department of Department of Forensic Medicine & Toxicology , Noida international University, Greater Noida, Uttar Pradesh 203201, India., hemantyadav.d@gmail.com

[4]Professor, Department of Pharmacology, Noida Institute of Engineering and Technology (Pharmacy Institute), Greater Noida, Uttar Pradesh, India, directorpharmacy@niet.co.in, 0000-0002-3053-8106

[5]Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India. himanshu.makhija.orp@chitkara.edu.in https://orcid.org/0009-0002-0864-9069

[6]Professor, Department of Orthopaedics, Vinayaka Mission's Kirupananda Variyar Medical College & Hospitals Vinayaka Mission's Research Foundation (DU), Tamil Nadu, India. ksoundar25@gmail.com, 0000-0001-7364-6732

[7]Associate Professor, Department of Bioinformatics, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India, alexanand.bioinfo@sathyabama.ac.in, 0000-0002-5842-5365

[8]DSc, Associate Professor, Department of Public Health and Healthcare Management, Tashkent State Medical University, Uzbekistan natyza@mail.ru, 0000-0003-3353-6031

ABSTRACT: Plant-derived natural products remain one of the most important sources of bioactive compounds for drug discovery owing to their remarkable chemical diversity and broad spectrum of pharmacological activities. However, conventional approaches for identifying therapeutically relevant phytochemicals are labor-intensive, time-consuming, and associated with high costs and low success rates during lead optimization. Recent advances in artificial intelligence (AI) have transformed natural product research by enabling rapid prediction of pharmacological activities through the integration of machine learning, deep learning, and cheminformatics approaches. AI algorithms can analyze large-scale chemical and biological datasets, identify structure–activity relationships, predict molecular targets, estimate pharmacokinetic and toxicity profiles, and prioritize promising compounds for experimental validation. The availability of publicly accessible phytochemical and bioactivity databases has further accelerated the development of robust predictive models for natural product-based drug discovery. This review summarizes recent advances in AI-assisted prediction of pharmacological activities of

plant-derived natural products, with particular emphasis on machine learning algorithms, deep learning architectures, molecular representations, and publicly available databases used for predictive modeling. Furthermore, the review discusses recent therapeutic applications across major disease areas, including cancer, infectious diseases, inflammatory disorders, diabetes, and neurodegenerative diseases, while highlighting current challenges, methodological limitations, and emerging opportunities for integrating explainable AI and multi-omics technologies into phytopharmaceutical research. Collectively, AI-driven computational strategies have the potential to accelerate natural product drug discovery, improve prediction accuracy, reduce experimental costs, and facilitate the development of safer and more effective plant-derived therapeutics.

1. INTRODUCTION

Natural products derived from plants have always been a major source of drugs and remain vital sources in today's drug discovery efforts. This extraordinary structural diversity, stereochemical complexity and biological specificity have led to the design of many clinically useful drugs for the treatment of cancer, infectious diseases, cardiovascular diseases, metabolic syndromes and neurological diseases (1). While many synthetic compounds do not have such a history, phytochemicals have been selected over the course of evolution to interact with a biological system and are thus useful as scaffolds for the discovery of novel drug candidates. The various types of plant secondary metabolites, such as alkaloids, flavonoids, terpenoids, phenolic compounds, glycosides and saponins, show a wide range of pharmacological activities that are still providing inspiration for pharmaceutical research and development (2). While all of these are benefits, the discovery of pharmacologically active natural products is still a time-consuming and difficult process. The conventional method of drug discovery involves a series of steps, including ethnopharmacological selection, extraction, isolation, purification, structural elucidation, biological screening, toxicity assessments, and full preclinical and clinical testing. These processes are labour-intensive, technologically demanding and very costly, and the chance that a bioactive compound will be successfully translated into an approved therapeutic product is still relatively low (3). Furthermore, there are thousands of chemically diverse constituents in medicinal plants, many of which have not yet been studied in detail and cannot be screened experimentally in a comprehensive way, without consuming too many resources. This increasing complexity of chemical structure has made it necessary to develop new computational approaches that can rank compounds to identify those that are most likely to have useful properties before they are tested in the lab. Artificial intelligence (AI) has proven itself as a game-changer to overcome numerous of these challenges and speed up several steps of the drug discovery process. AI can rapidly analyze large amounts of chemical and biological data that would be hard to interpret with traditional computational analysis, by applying techniques such as machine learning, deep learning, natural language processing and advanced data mining (4). AI models can help predict pharmacological activity, molecular target, toxicity profile, pharmacokinetic properties, and drug-likeness, thereby building increasingly accurate molecular-to-biological models that can reveal complex relationships between molecular structures and biological responses. This has made AI a valuable tool for the field of computational drug discovery, cutting the time and expense required for initial lead identification. With the advent of AI in cheminformatics, bioinformatics, and high throughput technologies, the opportunities of natural product research have grown even further. Sophisticated predictive algorithms have been built with the help of publicly available databases with chemical structures, bioactivity data, genomic data and pharmacological data. Machine learning techniques like Random Forest, Support Vector Machine, Decision Tree and Extreme Gradient Boosting, along with deep learning models like Graph Neural Networks and Convolutional Neural Networks, have yielded promising results in predicting the biological activity of plant-derived compounds and extracting potential therapeutic leads for experimental testing (5). These computational methods not only help make the screening process more efficient, but they also enable the identification of new bioactive molecules that would not be discoverable through traditional experimental screening pipelines. While there has been great progress, there are still some obstacles that stand in the way of the widespread adoption of AI in phytopharmaceutical research. The availability of high-quality well-annotated datasets is a key driver of performance for models, and the challenges of data heterogeneity, algorithms interpretability, data reproducibility and limited experimental validation are important hurdles for meaningful prediction. In addition, the properties of natural products are extremely variable, making their model generalization challenging from one therapeutic use to another. Closer integration of the skills of these four communities will be needed to

create clear and meaningful predictive models that can aid in evidence-based natural product drug discovery (6). In this review, the authors provide a detailed report on the recent progress in the methodology, predictive algorithms, databases accessible to the public, and applications of the AI-based approach in predicting the pharmacological activities of plant-derived natural products. Furthermore, it explores the present challenges, methodological limitations, and future trajectories that will shape the next generation of AI-driven phytopharmaceutical research, underscoring the increasing promise of computational intelligence to speed up the discovery and development of safe, effective, and innovative plant-derived therapeutics.

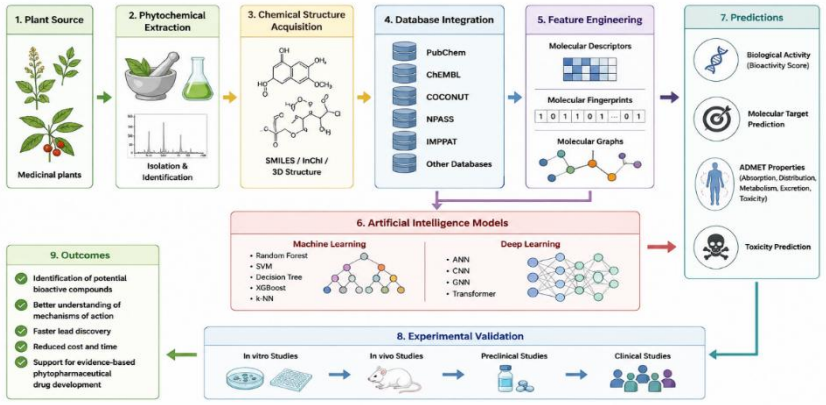

Figure 1. Artificial Intelligence Workflow for Predicting Pharmacological Activities of Plant-Derived Natural Products

2. ARTIFICIAL INTELLIGENCE IN NATURAL PRODUCT DRUG DISCOVERY

Artificial intelligence (AI) has proven to be one of the most impactful technologies in drug discovery, revolutionizing the process of discovering, optimizing, and developing bioactive molecules as potential drugs. Both traditional computational approaches, which are often based on known algorithms and statistical analyses, and AI approaches use data-driven learning techniques that can identify complex patterns in large, multidimensional datasets. The use of AI in natural product drug discovery has been a game-changer, allowing researchers to rapidly predict biological activities, molecular targets, toxicity, and pharmacokinetic properties of phytochemicals prior to experimental testing and validation (7). As a result, AI is now proving to be a game-changer in lowering both the time and the attrition rates of the conventional drug discovery process, and also the costs. High quality chemical and biological data from phytochemical studies, bioassays, genomics, metabolomics, and high throughput screening experiments are invaluable for the successful implementation of AI into natural product research. The datasets are preprocessed to eliminate redundant information, correct missing values, normalize chemical structures and inconsistencies that could affect model performance. Compounds are then converted into a computational-interpretable format, such as molecular descriptors, chemical fingerprints, the Simplified Molecular Input Line Entry System (SMILES) notation, molecular graphs, or a 3-D structural representation, after being curated. These molecular representations convey quantitative data of physicochemical properties and structural characteristics that aid the relationship between chemical structures and biological responses by the use of AI algorithms (8). In the field of natural product drug discovery, machine learning (ML) is one of the most widely used fields of AI. In ML algorithms, the patterns to be learned are not explicitly programmed but are learned directly from previously characterized datasets. Supervised learning techniques are used in prediction of properties of newly discovered phytochemicals from labeled sets of biological activity and unsupervised learning techniques discover hidden patterns, clusters, and structural similarity between plant compounds without specific outcome labels. Semi-supervised learning integrates both, and is especially useful in the case of limited natural product datasets that have been experimentally validated. A computational approach can be used to identify compounds, which can be prioritized for biological testing to reduce the number of unnecessary experiments conducted (9). The supervised learning algorithms like Random Forest, Support Vector Machine, Decision Tree, kNN, Naïve Bayes and XGB models have shown great success in predicting the pharmacological activities of plant-derived compounds. These models have been used to predict anticancer, antimicrobial, anti-inflammatory, antioxidant and antidiabetic activities, and have been employed to catch complex relationships between molecular descriptors and experimentally validated biological responses. They have been able to handle non-linear data, and have been able to find influential structural features that many traditional statistical methods have not been able to. In addition, ensemble learning methods that integrate several algorithms often result in more robust and generalizable models,

especially when dealing with chemically diverse phytochemical data sets (10). The recent developments of deep learning (DL) have also enhanced the potential of artificial intelligence (AI) in phytopharmaceutical research, as automated knowledge extraction from complex molecular data becomes possible. Whereas machine learning approaches rely on handcrafted descriptors to encode molecular information, the architectures of deep learning learn a hierarchy of molecular representations directly from the input data. The ability of Convolutional Neural Networks (CNNs) to recognize spatial molecular patterns and Graph Neural Networks (GNNs) to model atomic graph of chemical structures with more biologically meaningful representations of molecular interactions have proven successful. Transformer-based architectures have also been applied to molecular sequence analysis in natural language processing to predict molecular properties and biological activities for a variety of natural product libraries (11), with their use in molecular sequence analysis having been adopted recently using SMILES representations. The AI-driven drug discovery workflow typically involves several stages, including data acquisition and curation, molecular representation, feature extraction, model training, validation, and prediction. Performances of developed models are commonly assessed through statistical metrics like accuracy, precision, recall, F1 score, receiver operating characteristic (ROC) curve, area under the ROC curve (AUC), Matthews correlation coefficient and cross validation methods. The external validation with other sets of data further increases the reliability of models and confidence in the predicted pharmacological activities. The power of computation, the sophistication of the algorithms, and the amount of publicly available data is growing, and AI has increasingly been able to detect novel bioactive phytochemicals that can't be found by screening methods. Thus, AI's ability to combine medicinal chemistry, pharmacology, and systems biology is creating a more efficient and rational pathway towards plant-derived natural product drug discovery and bringing the leap from computational prediction to experimental validation and therapeutics one step closer (12).

Table 1. Comparison of Artificial Intelligence Algorithms Used for Predicting Pharmacological Activities of Plant-Derived Natural Products

AI Technique	Learning Type	Key Applications	Strengths	Limitations
Random Forest (RF)	Supervised	Bioactivity, QSAR	High accuracy; handles complex and high-dimensional data	Less interpretable than simpler models
Support Vector Machine (SVM)	Supervised	Classification, target prediction	Effective for small and nonlinear datasets	Sensitive to kernel selection and parameter tuning
Decision Tree (DT)	Supervised	Classification	Simple, interpretable, and easy to implement	Prone to overfitting
XGBoost	Supervised	Bioactivity prediction, ADMET prediction	High predictive performance and computational efficiency	Requires careful hyperparameter tuning
k-Nearest Neighbors (k-NN)	Supervised	Similarity analysis, classification	Simple algorithm with no explicit training phase	Performance decreases with large datasets
Artificial Neural Network (ANN)	Deep Learning	QSAR, molecular property prediction	Captures complex nonlinear relationships	Requires large datasets and longer training time

3. AI-BASED PREDICTION OF PHARMACOLOGICAL ACTIVITIES

3.1 Molecular Representation and Feature Engineering

The accuracy of the chemical compound representation before model development is a key factor in the predictive ability of the AI models. The structural complexity of phytochemicals is not easily understood by machine learning algorithms, so molecular structures need to be converted into some numerical representation that conveys their physicochemical and structural properties. The process of molecular representation or feature engineering is the basis for AI-assisted pharmacological activity

prediction (13). Molecular representations that are effective help computational models identify patterns between molecules and biological responses, which will lead to higher prediction accuracy and better identification of new bioactive compounds. A number of computational methods have been devised for representation of natural products of plant origin. Molecular descriptors are quantitative data related to molecular mass, lipophilicity, hydrogen bonding donors and acceptors, topological indices, electronic properties and three-dimensional conformation. These descriptors are a succinct representation of compounds' physicochemical properties and are commonly used in quantitative structure–activity relationships (QSAR). Furthermore, molecular fingerprints can be used to transform the chemical structure into a binary or numerical vector which identifies the presence or absence of particular features of the structure or function. These popular fingerprinting methods, such as Extended Connectivity Fingerprints (ECFP), Molecular ACCess System (MACCS) keys and Morgan fingerprints, allow for similarity analysis and model training in large phytochemical data sets in a short time window (14). In recent years, graphical molecular representations that model compounds as a network of connected atoms and chemical bonds have been introduced thanks to the progress of deep learning. As opposed to classic descriptor-based approaches, graph representations retain the original topology of molecular structures and enable AI algorithms to learn structure information directly from the molecular graph. It is especially beneficial for plant-based natural products, which often contain complex ring structures, stereochemical features, and a variety of functional groups that cannot be fully described with conventional descriptors alone. Moreover, the molecular representation called Simplified Molecular Input Line Entry System (SMILES) has been adopted as a standard convention for the representation of chemical structures, facilitating the processing of chemical structures by transformer-based deep learning models in a sequence-based format. The complementary representation strategies enable a comprehensive structural information which enables the AI models to predict pharmacological activities of various phytochemical classes (15). Effective preprocessing techniques are also crucial when it comes to the quality of feature engineering. This can lead to poor predictive performance, and the risk of overfitting increases when there are redundant descriptors, missing values, variables that are highly correlated, and noisy data. Many feature selection algorithms have been used to find the most informative variables and to remove the uninformative ones to reduce the computational needs, including recursive feature elimination (RFE), principal component analysis (PCA), and mutual information analysis. Feature optimization is also a technique that not only helps to improve the efficiency of a model but also makes it more interpretable and reproducible when using natural products of plant origin to create a pharmacological prediction model.

3.2 Machine Learning-Based Prediction of Pharmacological Activities

Due to its capacity to learn complex relationships between molecular properties and experimentally observed pharmacological effects, machine learning is one of the most widely used approaches in AI for the prediction of the pharmacological activities of natural products. Supervised learning algorithms are especially useful in the field of phytopharmaceutical research as they are trained with compounds that have known pharmacological effects and can then predict the biological activity of previously uncharacterized compounds. These algorithms allow researchers to quickly narrow their search to potentially fruitful lead compounds by examining hidden nonlinear relationships in multidimensional datasets, without having to go to the laboratory and test all the possibilities (16). Random Forest is one of the most frequently used algorithms among the several machine learning algorithms, as it is robust, not prone to overfitting, and can operate on high dimensional data sets. The algorithm builds several decision trees from randomly sampled subsets of data and fuses the predictions from these decision trees to boost classification and regression accuracy. The current study has shown that Random Forest is a powerful tool for the prediction of anticancer, antimicrobial, anti-inflammatory, antioxidant and antidiabetic activity of plant extracts and relative importance estimation of molecular descriptors. Support Vector Machine is another algorithm that is widely accepted due to its capability to accurately classify a complex nonlinear dataset by kernel based learning. This algorithm is especially useful in the case of relatively small training sets, which is often encountered in natural product research. Other machine learning algorithms have been added to computational pharmacology' predictive prowess such as Decision Tree, k-Nearest Neighbors, Naïve Bayes, Logistic Regression, and Extreme Gradient Boosting (XGBoost). One of the reasons for the huge attention of XGBoost is its computational efficiency, scalability, and the capacity to deal with missing values and minimize the prediction errors by gradient boosting optimization. Combining the strengths of several models using ensemble techniques can yield more reliable predictions, reduce bias and instability, while also offering greater accuracy. Selecting a suitable machine learning algorithm has been shown to be dependent of the size of the dataset, molecular diversity, the type of chemical descriptors used, and the drug pharmacologic endpoint rather than on a single best model in comparative evaluations (17). Machine learning models have become a standard tool to predict several pharmacological properties at once, such as biological activity, target specificity,

toxicity, absorption, metabolism, distribution and drug-likeness. The combination of these predictive capabilities into an integrated computational framework has greatly enabled scientists to shorten the time required for early drug discovery through less experimental work, and more efficient prioritization of phytochemicals for their further validation in vitro and in vivo. The use of Machine Learning is anticipated to be one of the most impactful computational technologies for the discovery of plant-derived therapeutic agents, as increasingly diverse and high-quality datasets become available.

3.3 Deep Learning-Based Prediction of Pharmacological Activities

The advent of deep learning has greatly improved the ability of AI models to forecast natural products' pharmacological properties, offering promising prospects for discovering novel drug candidates from the plant kingdom.The insights and advancements gained from deep learning have greatly improved the ability to predict natural products' pharmacological activities, paving the way for the discovery of novel drug candidates from the plant kingdom. Conventional machine learning algorithms typically rely on the manually chosen molecular descriptors and feature engineering. Deep learning models automatically learn the hierarchy of the features representations directly from the raw chemical data. This function can be used to identify subtle structural properties and nonlinear relationships that might otherwise be difficult to identify through conventional computational techniques. Consequently, deep learning has become more and more useful in tackling the structural diversity and chemical complexity of natural products, which has also enhanced the predictions of their bioactivity and enabled the rapid identification of lead compounds (18). One of the following deep learning architectures has been successfully applied by extracting complex structural patterns from molecular images and descriptor matrices for predicting molecular properties: Convolutional Neural Networks (CNNs). CNNs are especially useful in learning about spatial relationships between molecular features, which could be used to predict bioactivity, toxicity, and physicochemical properties of phytochemicals. Recurrent Neural Networks (RNNs), which were first used for sequential data analysis, have also been used to analyze sequences of molecules in the form of Simplified Molecular Input Line Entry System (SMILES) notation. More recently, transformer architecture has shown to provide greater accuracy in learning long-range structural dependencies in molecular sequences, giving greater predictive power of the biological activity and molecular interactions in a wide range of phytochemical libraries. One of the most significant breakthroughs in AI-enabled molecular modelling is the advent of Graph Neural Networks (GNNs), which harness the graph-like structure of chemical compounds. The GNNs share the same structure as networks used in machine translation, where atoms are treated as nodes and chemical bonds as connections.Unlike machine translation networks, the GNNs keep the topology of molecules intact when learning features. This representation is useful for natural products, which often feature complex ring systems, stereochemical differences, and several functional groups, making it challenging to represent such molecules with conventional descriptor-based approaches. As a result, GNNs have shown great promise in predicting molecular properties, protein–ligand binding, target affinity, and even pharmacological activities, thereby alleviating the loss of information provided by conventional feature extraction approaches (19). Deep learning has also enabled multitask learning, transfer learning and self-supervised learning, which further enhance the predictive accuracy in the absence of experimental validation datasets. Unlike single-task learning which models several biological targets with distinct molecular representations, multi-task learning is able to predict several biological targets with shared molecular representations, while transfer learning allows knowledge to be transferred from large chemical datasets to smaller natural product datasets. Self-supervised learning then minimizes reliance on labeled data by enabling neural networks to learn useful molecular representations from unlabeled molecules and then fine-tune for a specific prediction task. All of these high-level learning strategies have broadened the scope of deep learning's use in phytopharmaceutical research, and they are making drug discovery more efficient. While deep learning models are often more powerful in their predictive abilities, they typically need more data and computational power than traditional machine learning models. Optimization, hyperparameter tuning, and training complexity are practical issues that often arise in model optimization, especially when high quality natural product datasets are scarce. However, the ever-evolving capabilities of computational technology, cloud computing, GPUs and access to molecular databases are increasingly making deep learning more accessible for natural product research, thereby expanding its application in the discovery of novel therapeutics from plants.

3.4 Prediction Across Major Therapeutic Areas

AI has been used with success and is able to predict the pharmacological activities of natural products from plant source in a wide variety of therapeutic areas, offering researchers efficient strategies to identify promising lead compounds prior to experimental validation. Machine learning and deep learning models have been used extensively in cancer research to discover

phytochemicals that possess anti-proliferative, pro-apoptotic, antiangiogenic, and metastasis-inhibiting properties. A key benefit of AI is its ability to combine molecular descriptors, transcriptomic profiles, and bioactivity data to prioritize compounds for in-lab testing that demonstrate strong anticancer activity, thus minimizing the number of candidates that need to be tested in the laboratory (20). AI is proving to be a valuable asset in antimicrobial drug discovery, especially in the context of combating antimicrobial resistance (AMR) – a worldwide issue that demands new solutions. Predictive models developed from chemical and microbiological data can be used to quickly identify compound(s) found in plant extracts with antibacterial, antifungal, antiviral and antiparasitic properties. Extensive use of computational screening has led to the identification of molecules that could be used to target resistant pathogens and has greatly reduced the time needed for the conventional bioassay-guided screening. This same strategy has been applied to the prediction of interactions with signaling pathways, inflammatory mediators, cytokines and enzyme targets in chronic inflammatory disorders to identify natural products with anti-inflammatory properties. AI also has shown significant promise in predicting antidiabetic and neuroprotective properties of medicinal plant constituents. The use of machine learning algorithms to discover compounds with the ability to modulate glucose metabolism, insulin signaling pathways, oxidative stress, and inflammatory responses involved in diabetes mellitus is an increasing trend. Similarly, using AI to screen compound libraries has helped identify phytochemicals which bind to enzymes, receptors and signaling pathways that are linked to neurodegenerative diseases like Alzheimer's disease and Parkinson's disease. The combination of molecular docking, network pharmacology, and the application of AI-driven predictive modeling has given scientists more confidence in their computational predictions and given them valuable insights into the mechanisms at play prior to experimental validation. In addition to forecasting single-target pharmacology, modern AI platforms are now able to assess a drug's effect on multiple targets, as well as its efficacy, toxicity, and pharmacokinetic profile. These comprehensive prediction models are especially helpful for natural products since many phytochemicals may act through several molecular targets than just one biological target. The integration of cheminformatics, systems biology, and advanced AI algorithms can help researchers gain a deeper understanding of the intricate relationships among phytochemicals and biological systems, which can help increase the success rate of selecting promising candidates for preclinical investigation. The prediction of pharmaceutical activities of natural products from plant sources will be more accurate, interpretable, and clinically relevant with the continuous evolution of AI methodologies and the growing number of chemical and biological data (21,22).

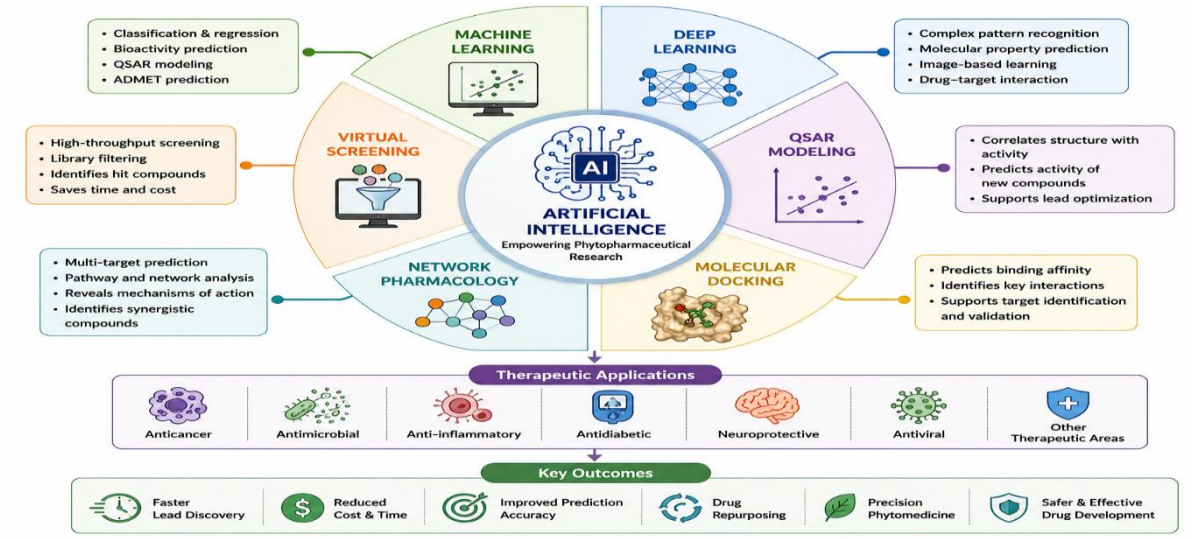

Figure 2. Applications of Artificial Intelligence in Plant-Derived Natural Product Drug Discovery

4. DATABASES AND RECENT APPLICATIONS

igh-quality information in the form of complete chemical and biological databases have been instrumental in the fast progress of AI in natural product drug discovery. They feature curated records of molecular structure, physicochemical properties, biological activities, targets of protein interaction, metabolic pathways, pharmacokinetic properties and toxicities, which enables building of powerful machine learning and deep learning models to predict pharmacological activity. These continually growing public databases have further enhanced the confidence in computational models as they provide a greater diversity of data and

promote reproducible research. Additionally, the implementation of standardized data formats and the regular updates of databases have improved the interoperability between cheminformatics, bioinformatics, and pharmacological platforms, enabling the easy integration of heterogeneous data into AI-assisted drug discovery processes (23). One of the most popular resources is PubChem, which is one of the largest public repositories for chemical compounds, bioassay data, and molecular properties, making it an indispensable resource for using AI to predict biological activities. ChEMBL contains manually curated information on bioactive molecules, experimentally validated pharmacological data, target proteins and drug like compounds, which is used to develop predictive models for target identification and activity estimation. With the advent of databases especially devoted to natural products, such as COCONUT (Collection of Open Natural Products), NPASS (Natural Product Activity and Species Source Database), and IMPPAT (Indian Medicinal Plants, Phytochemistry and Therapeutics Database), these databases have proved to be extremely useful since they combine the phytochemical data with the biologically active components and their plant sources that have been reported in experiments. These resources can be used to analyze structurally diverse phytochemicals and to enhance prediction of therapeutic properties in many disease areas (24). There are also some computational platforms and open-source software packages that have helped implement the use of AI algorithms in natural product research. Molecular descriptor calculation and generation of fingerprints, feature engineering, model construction and virtual screening can all be performed in a comprehensive environment using tools like RDKit, DeepChem, KNIME, AutoQSAR, and PASS Online. These platforms can be coupled with publicly available databases, enabling researchers to quickly preprocess chemical data, compare various machine learning algorithms, assess predictive performance, and select promising lead compounds for further experimental study. This computational power has significantly lowered the technical hurdles of implementing AI, paving the way for its wider application in pharmaceutical and academic settings to support predictive modeling. In the last ten years, the use of AI in forecasting pharmacological activities has grown significantly. Many studies have shown that machine learning algorithms can accurately classify phytochemicals based on their biological activities, prioritize compounds with suitable pharmacokinetic properties, and discover new molecular targets linked to the disease progression. By minimizing the need for experimental testing of compounds, AI-driven virtual screening has streamlined the initial phase of drug discovery, boosting efficiency and conserving valuable resources. Moreover, the molecular docking, network pharmacology and AI prediction has helped in exploring the complex mechanisms of action and understanding multitarget interactions that are often found with plant-derived natural products (25). In recent years, artificial intelligence has been used in conjunction with multi-omics technologies (such as genomics, transcriptomics, proteomics, and metabolomics) to deepen the understanding of disease mechanisms and therapeutic responses. Combining these complementary data sets, AI models can find biomarkers, anticipate target–compound interactions, and create extensive molecular networks to aid precision phytopharmaceutical studies. Integrative strategies are especially useful in complex diseases like cancer, diabetes, neurodegenerative disorders and chronic inflammatory diseases, where there are multiple pathways involved in the disease process. The increasing use of explainable AI techniques has even further increased confidence in computational predictions by making models transparent by making them explainable by pharmacologists and medicinal chemists with regard to the molecular features that lead to a model decision. Nevertheless, the effectiveness of AI-driven drug predictions in medicine hinges on the quality and comprehensiveness of drug databases. There is a need for continuous curation, standardisation of annotations, experimental validation, and international data-sharing efforts to enhance the confidence in predictive models and ensure their maximum translation. With continued growth in the size of the database and the ongoing development of computational methods, AI-assisted natural product research will increasingly be able to identify bioactive phytochemicals faster and facilitate the creation of new therapeutic agents. Utilizing all-encompassing databases, cutting-edge computational resources, and cross-disciplinary collaboration will further empower AI to expedite the process of drug development from plants and support evidence-based phytopharmaceutical research (26).

5. CURRENT CHALLENGES

While AI has made significant strides in natural product drug discovery, there are still several challenges that must be overcome before it can be fully integrated into the clinical process and become a standard tool for drug discovery.Although AI has already made significant strides in the field of natural product drug discovery, there are still several challenges that need to be addressed before it can be fully embraced in the clinical process and become a routine tool for drug discovery. The scarcity of high quality, standardized, and experimentally validated datasets is one of the most important issues. Large amounts of reliable data are essential for the development of strong predictive models by AI algorithms; however, data pertaining to plant-derived natural products is scattered in many databases and collected with different experimental approaches, and annotations are not always

consistent. This variation makes it difficult to share data and can impact the validity, repeatability and applicability of predictive models. Moreover, many of the medicinal plants and the different phytochemicals present in them are poorly described and characterized, leading to incomplete data which limits the creation of comprehensive prediction systems based on artificial intelligence (27). Another interesting challenge is the tremendous structural variety and complexity of the naturally occurring compounds present in plants. Phytochemicals often contain complex stereochemistry and multiple chiral centers, unique ring systems and highly diverse functional groups, which is difficult to represent through traditional molecular descriptors and is not found in synthetic chemical libraries. While recent developments in graph-based learning and transformer architectures have brought molecular representation a step further, the accurate representation of the characteristics of the molecule that leads to biological activity is still a great computational challenge. Moreover, bringing together heterogeneous data from genomics, transcriptomics, proteomics, metabolomics and pharmacology will require complex data harmonization approaches to ensure interoperability between these various data sources. Another challenge in AI-powered drug prediction is model interpretability. While many advanced machine learning and deep learning algorithms can deliver high accuracy predictive models, they are "black-box" models with little biological understanding of the reasons and rationale behind their decisions. Lack of transparency diminishes trust in the process among researchers and regulators, especially during compound prioritisation based on AI predictions for experimental validation or clinical development. Given these reasons, a growing focus has shifted to methods that can make the AI model more transparent, but maintain its accuracy.As a result, there is a growing interest in "explainable" approaches to AI that can make the model more transparent while retaining its predictive ability. Beyond the technical challenges, computational resources and the combined expertise of multiple fields are indispensable for effective application of AI. For the development of reliable predictive models, there is a need to coordinate the work of pharmacologists, medicinal chemists, computational biologists, bioinformaticians and data scientists to make sure that the data are suitably prepared, the models are selected and validated, and the biological interpretation is correct. Solving these scientific, technical and collaborative problems will be essential to enhancing the reliability, reproducibility and applicability of AI-driven prediction models and enable them to be incorporated into evidence-based phytopharmaceutical drug discovery (28).

6. LIMITATIONS

While artificial intelligence shows great promise in predicting the pharmacological activities of natural products from plants, there are a few inherent limitations that still limit its use in the field of drug discovery. One of the main hurdles is the reliance of AI models on existing experimental data, which are scarce and not uniformly available across all therapeutic areas, and skewed toward more well-studied phytochemicals. As such, predictions might yield very high accuracy for compounds similar to those used to train the model, and less accuracy for structurally novel or chemically diverse natural products. This restriction is especially significant for species that have been little studied scientifically to date, and for which the extent of biological research is not sufficient to support development of a model and testing it against the external environment. Another potential constraint is the ability to transfer predictive models to different biological systems and disease states. Predictive models produced with specific data sets or pharmacological endpoints might not retain predictive ability in other independent data sets when experiments are performed under different conditions. Furthermore, inconsistencies can be present due to differences in extraction techniques, phytochemical composition, assay procedures, and biological models, which can limit the reliability of AI-predictions. These sophisticated deep learning models are also computationally intensive, requiring more powerful computing hardware and specialized skills, which can make their use challenging in resource-limited research environments. Also, theoretical predictions do not replace experimental validation. Hypotheses developed by AI must be ultimately tested in vitro, in vivo and clinically to demonstrate efficacy, safety, pharmacokinetic properties, and therapeutic relevance. Thus, AI should be seen as an aid to decision making for enhancing the effectiveness of natural product research, not as a replacement for drug testing studies. To break these barriers and achieve the full potential of AI-driven phytopharmaceutical discovery, there are several areas that require further development and refinement of data quality, model transparency, external validation, and interdisciplinary collaboration (29).

7. FUTURE PERSPECTIVES

The field of natural product research is poised to increasingly leverage artificial intelligence in the future, as the integration of tools and methods in computational sciences, biological data generation, and collaborative efforts among disciplines are driving the evolution of modern drug discovery. The increasing number of high-quality phytochemical, pharmacological, genomic and metabolomic datasets will provide the basis for developing more reliable and transferable predictive models to identify new

bioactive compounds more effectively. The next generation of AI systems are expected to combine various data modalities into single computational frameworks that will allow for the understanding of intricate relationships among molecular structures, biological pathways, disease mechanisms, and therapeutic responses (30). One of the most promising developments is the emergence of explainable artificial intelligence (XAI), which aims to improve the transparency and interpretability of predictive models. With its ability to offer clear explanations for computational predictions, XAI can improve the trust in the prediction by researchers, clinic and regulatory agencies and ease the identification of molecular features responsible for the pharmacological activity. The ability of the models to be interpreted will further augment the scientific validation as well as rational decision making in lead optimization and preclinical drug development. At the same time, generative artificial intelligence is anticipated to revolutionize the identification of structurally novel phytochemical analogues, by creating molecules that are optimized for biological activity but retain the desirable pharmacological properties of their parent compounds. Phytopharmaceutical research is poised to reach even more significant heights with the integration of AI, systems biology, network pharmacology, molecular docking, and multi-omics technologies. The complementary methods will allow simultaneous assessment of various biological targets and signaling pathways, which is representative of the multitarget activity of many compounds of plant origin. These integrative approaches could help uncover combination therapy, identify synergistic phytochemical interactions and create individualized phytomedicine based on the genetic and molecular characteristics of each person. Furthermore, the advent of cloud computing, high-performance computing, and open-access scientific databases will enable researchers around the world to increasingly implement sophisticated AI methods, fostering collaborative innovation and facilitating reproducibility between different institutions. The future also relies on the adoption of a standardised data-reporting protocol, on-going database curation, international sharing of data, and enhanced partnerships between medicinal chemists, pharmacologists, computational biologists, bioinformaticians and experts in artificial intelligence. All these will help to improve the translatability, interpretability, and reliability of AI-driven drug prediction in medicine, thereby speeding up the discovery of safe, effective, and clinically relevant plant-based therapeutics, and aiding the progression of evidence-based phytopharmaceutical drug discovery.

8. CONCLUSION

Artificial intelligence (AI) has become a game-changing technology that is helping to transform the discovery and development of natural products from plants by predicting their pharmacological activities quickly, cost-effectively, and with data. The use of machine learning, deep learning, cheminformatics and bioinformatics has significantly enhanced the identification of potential phytochemicals, with accurate prediction of biological activity, molecular targets, toxicity, pharmacokinetic properties and drug-likeness. Combined with the growing number of curated phytochemical databases and powerful computational tools, AI has fast-tracked the early screening process of drug discovery, thereby minimizing the need for comprehensive experimental screening. Despite the progress made, there are still several scientific and technical challenges that hinder the potential of AI in phytopharmaceutical research to its fullest extent.While these are significant achievements, there remain a number of scientific and technical challenges that need to be addressed for the full potential of AI in phytopharmaceutical research to be realised. These challenges reflect the need for further developments in computational methods and biological data to improve the ability to overcome these limitations. To overcome these challenges, there is a need for standardized data collection, standardized model validation, and multidisciplinary collaboration, which will further improve the reliability and translatability of AI-based prediction systems. In the future, the use of explainable AI, generative models, network pharmacology, molecular docking, and multi-omics will further bolster computational drug discovery and insight into the intricate mechanisms of medicinal plant therapeutic action. These improvements will help to identify new bioactive substances, aid in the discovery of new drugs with multiple targets and contribute to the advancement of precision phytomedicine. In conclusion, AI should not be seen as a substitute for experimental pharmacology, but rather as a valuable tool that can complement it in various ways, such as supporting decision-making, improving resource efficiency, and speeding up the identification of safe and effective plant-based therapeutics with clinical relevance. To realize these benefits, it will be essential for the fields of computational scientists, pharmacologists, medicinal chemists and experimental researchers to work closely together to bring the AI-driven predictions into the realm of validated therapeutic innovations that can have a significant impact on future healthcare.

DECLARATIONS

Funding: The authors received no specific financial support for the research, authorship, and/or publication of this article.

Conflict of Interest: The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Author Contributions: All authors contributed substantially to the conception, literature review, manuscript preparation, critical revision, and approval of the final version of the manuscript.

Ethical Approval: Ethical approval was not required because this study is a review article based exclusively on published literature.

REFERENCES

[1] Mullowney MW, Duncan KR, Elsayed SS, Garg N, van der Hooft JJJ, Martin NI, et al. Artificial intelligence for natural product drug discovery. Nat Rev Drug Discov. 2023;22(11):895-916. doi:10.1038/s41573-023-00774-7.
[2] Saldívar-González FI, Aldas-Bulos VD, Medina-Franco JL, Plisson F. Natural product drug discovery in the artificial intelligence era. Chem Sci. 2022;13:1526-1546. doi:10.1039/D1SC04471K.
[3] Tropsha A, Isayev O, Varnek A, Schneider G, Cherkasov A, et al. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. Nat Rev Drug Discov. 2024;23:141-155.
[4] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. 2012;40(Database issue):D1100-D1107.
[5] Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 2019;47(D1):D930-D940.
[6] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem in 2021: new data content and improved web interfaces. Nucleic Acids Res. 2021;49(D1):D1388-D1395.
[7] Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: collection of open natural products database. J Cheminform. 2021;13:2.
[8] Zeng X, Zhang P, Wang Y, Qin C, Chen S, He W, et al. NPASS: natural product activity and species source database. Nucleic Acids Res. 2018;46(D1):D1217-D1222.
[9] Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. Curr Protoc. 2021;1:e90.
[10] Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, Samal A. IMPPAT: a curated database of Indian medicinal plants, phytochemistry and therapeutics. Sci Rep. 2018;8:4329.
[11] Newman DJ, Cragg GM. Natural products as sources of new drugs over nearly four decades. J Nat Prod. 2020;83(3):770-803.
[12] Atanasov AG, Zotchev SB, Dirsch VM, Supuran CT. Natural products in drug discovery: advances and opportunities. Nat Rev Drug Discov. 2021;20:200-216.
[13] Lo YC, Rensi SE, Torng W, Altman RB. Machine learning in chemoinformatics and drug discovery. Drug Discov Today. 2018;23(8):1538-1546.
[14] 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;18:463-477.
[15] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-1250.
[16] Walters WP, Murcko MA. Assessing the impact of generative AI on medicinal chemistry. Nat Biotechnol. 2020;38:143-145.
[17] Ekins S, Puhl AC, Zorn KM, Lane TR, Russo DP, Klein JJ, Hickey AJ, Clark AM. Exploiting machine learning for end-to-end drug discovery and development. Nat Mater. 2019;18:435-441.
[18] Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, et al. QSAR modeling: where have you been? Where are you going to? J Med Chem. 2014;57:4977-5010.
[19] Muratov EN, Bajorath J, Sheridan RP, Tetko IV, Filimonov D, et al. QSAR without borders. Chem Soc Rev. 2020;49:3525-3564.
[20] Stokes JM, Yang K, Swanson K, Jin W, Cubillos-Ruiz A, Donghia NM, et al. A deep learning approach to antibiotic discovery. Cell. 2020;180:688-702.
[21] Karibasappa C, Kiran R, Nagarathna N, Sreenivasa G. Seasonal study of dung beetles assemblage at Chitradurga, Karnataka, India. J Entomol Res. 2025;49:1225-1230. doi:10.5958/0974-4576.2025.00217.1.
[22] Khurana B, Dhanker R. Biotoxic impact of distillery effluent on broccoli and termites. J Entomol Res. 2025;49(4):987-991. doi:10.5958/0974-4576.2025.00170.2.

[23] Patil AS, Marwaha L, Singh R. Entomological assessment of synthetic polymer breakdown: comparative study of PET, PU and nylon degradation by greater wax moth (Galleria mellonella). J Entomol Res. 2025;49(2):463-470. doi:10.5958/0974-4576.2025.00077.3.
[24] Ravindranath N, Samanta A, Samanta S. Dose-dependent toxicity and field efficacy of PEG-emulsified botanical formulations against chickpea pod borer (Helicoverpa armigera). J Entomol Res. 2025;49(3):632-637. doi:10.5958/0974-4576.2025.00110.3.
[25] Nair N, Thangjam B. Biotic and abiotic stresses on Apis cerana colonies in Tripura, N.E. India. J Entomol Res. 2025;49(2):476-480. doi:10.5958/0974-4576.2025.00079.6.
[26] Issaka E, Amu-Darko JNO. Biomimetic nanoparticles for cancer therapy: a review of recent advances, applications, and bottlenecks. Biomed Mater Devices. 2025;3(1):193-215.
[27] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database. Nucleic Acids Res. 2018;46:D1074-D1082.
[28] Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, et al. ZINC20—a free ultralarge-scale chemical database for ligand discovery. J Chem Inf Model. 2020;60:6065-6073.
[29] Wang Y, Bryant SH, Cheng T, Wang J, Gindulyte A, Shoemaker BA, et al. PubChem BioAssay: 2022 update. Nucleic Acids Res. 2023;51:D1505-D1514.
[30] Fourches D, Muratov E, Tropsha A. Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling. J Chem Inf Model. 2010;50:1189-1204.